Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **428,** 2548–2564 (2013) doi:10.1093/mnras/sts221

Halo occupation distribution modelling of green valley galaxies

Elisabeth Krause,[1,2]★ Christopher M. Hirata,[1] Christopher Martin,[3] James D. Neill[3] and Ted K. Wyder[3]

[1]*Department of Astronomy, California Institute of Technology, MC 249-17, Pasadena, CA 91125, USA*
[2]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[3]*Department of Astronomy, California Institute of Technology, MC 290-17, Pasadena, CA 91125, USA*

Accepted 2012 October 15. Received 2012 October 16; in original form 2012 August 30

ABSTRACT

We present a clustering analysis of near-ultraviolet (NUV)–optical colour selected luminosity bin samples of green valley galaxies. These galaxy samples are constructed by matching the Sloan Digital Sky Survey Data Release 7 with the latest *Galaxy Evolution Explorer* source catalogue which provides NUV photometry. We present cross-correlation function measurements and determine the halo occupation distribution of green valley galaxies using a new multiple tracer analysis technique.

We extend the halo occupation formalism, which describes the relation between galaxies and halo mass in terms of the probability $P(N, M_{\rm h})$ that a halo of given mass $M_{\rm h}$ contains N galaxies, to model the cross-correlation function between a galaxy sample of interest and multiple tracer populations simultaneously. This method can be applied to commonly used luminosity threshold samples as well as to colour and luminosity bin selected galaxy samples, and improves the accuracy of clustering analyses for sparse galaxy populations.

We confirm the previously observed trend that red galaxies reside in more massive haloes and are more likely to be satellite galaxies than average galaxies of similar luminosity. While the change in central galaxy host mass as a function of colour is only weakly constrained, the satellite fraction and characteristic halo masses of green satellite galaxies are found to be intermediate between those of blue and red satellite galaxies.

Key words: galaxies: evolution – large-scale structure of Universe.

1 INTRODUCTION

Most nearby galaxies fall into one of two well-known and well-characterized categories. They are either passively evolving elliptical galaxies with old stellar populations, red in colour and typically living in high-density regions, or actively star-forming spiral galaxies with blue colour. The latter often are field galaxies or reside in other low-density regions like cluster outskirts.

This blue/red galaxy colour bimodality has been observed to be in place already around $z \sim 1$. The fraction of red galaxies increases with time (e.g. Faber et al. 2007) and therefore galaxies must transition from blue to red. Galaxies in this transitional stage characteristically show low levels of recent star formation. As ultraviolet emission is a sensitive tracer of recent star formation, these transition galaxies are easily identified in a $(\mathrm{NUV} - r) - M_r$ colour–magnitude diagram where they populate a 'green valley' between well-localized red and blue sequences (Wyder et al. 2007).

The relation between galaxy colour and environment density also evolves with redshift, such that the fraction of red galaxies increases with time in dense environments but stays nearly constant for field galaxies (e.g. Cooper et al. 2007, and references therein). This indicates the transition from blue to red galaxies may be driven by environmental processes, associated with the infall of a galaxy into a larger halo ('cluster'). Proposed mechanisms broadly fall into one of the following categories: galaxy–galaxy interactions, such as galaxy mergers, merger-driven nuclear activity and high-speed galaxy interactions, galaxy–intracluster medium interactions (e.g. ram pressure stripping or thermal evaporation) and interactions between an infalling galaxy and the cluster potential (e.g. truncation through tidal forces). Observationally these are disentangled through their characteristic time-scales, the dependence of their respective efficiencies on halo mass and position within the cluster (Treu et al. 2003; Cooper et al. 2006; Moran et al. 2007); for example, galaxy mergers are expected to be one of the dominant processes in group-scale haloes and in the outskirts of massive clusters.

In the framework of Λ cold dark matter (ΛCDM) cosmology, the evolution and spatial distribution of dark matter haloes is relatively well understood. A common technique for inferring the masses of haloes hosting different galaxy populations is to measure the angular or spatial clustering of galaxies and relate it to the predicted

★ E-mail: ekrause@astro.caltech.edu

clustering and abundance of dark matter haloes. While the relation between galaxy and dark matter clustering on large scales can be approximately described by scale-independent biasing, the situation is more complicated – and more informative about the physical processes at work – on small scales. At the level of individual haloes, the so-called halo occupation distribution (HOD) models (e.g. Berlind & Weinberg 2002) describe the relation between galaxies and mass in terms of the probability that a halo of given mass contains N galaxies of a given type. Then galaxy clustering, e.g. the two-point correlation function, is modelled as the sum of contributions from galaxy pairs residing in the same halo and from galaxy pairs living in different haloes.

This method of interpreting galaxy correlation functions has been used extensively. For example, Zehavi et al. (2011, see references therein for previous/high-z studies) analyse the completed Sloan Digital Sky Survey Data Release 7 (SDSS DR7), and find, in agreement with previous results, that the amplitude of the correlation function increases with luminosity, and that at fixed luminosity redder galaxies are more strongly clustered, due to redder galaxies being satellites in more massive (and thus more biased) haloes. Based on correlation function measurements over the redshift range $0.2 < z < 1.2$ from the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS), Coupon et al. (2011) also find red central galaxies to reside in more massive haloes than average central galaxies in the same luminosity sample.

The clustering of (NUV $-$ r) colour selected galaxies from the *Galaxy Evolution Explorer* (*GALEX*) survey has previously been studied by Heinis et al. (2007), who measure the angular correlation function; Heinis et al. (2009) and Loh et al. (2010) analyse spatial clustering as a function of star formation history and colour, respectively. These authors find the clustering of green galaxies to have intermediate strength compared to blue and red galaxies and to have a scale dependence closer to that of red galaxies. At small scales their analysis is strongly limited by statistics due to the small number density of green valley galaxies, limiting their ability to constrain the one-halo term.

We extend the HOD formalism to simultaneously model the cross-correlation functions (CCFs) of a sparse luminosity bin galaxy sample with multiple more abundant galaxy populations to study the environment of local green valley galaxies. We consider luminosity bin samples of green valley galaxies as the physical mechanisms populating the green valley, i.e. quenching star formation in blue galaxies or rejuvenating red galaxies may depend on halo mass and thus vary with galaxy luminosity. Compared to an autocorrelation function based clustering analysis, measuring the CCF between (sparse) *GALEX* selected galaxies and more abundant samples of SDSS galaxies reduces the shot noise contribution to our measurements, and also increases the effective volume probed beyond the combined *GALEX*–SDSS footprint.[1] Extending previous work on HOD models for CCFs (e.g. Krumpe, Miyaji & Coil 2010) to simultaneously fit the clustering of the galaxy sample of interest with respect to multiple tracer populations is particularly helpful for analysing the clustering of luminosity bin samples, which are harder to constrain than the more frequently used luminosity threshold samples. This allows us to put the separate piece of information found by Heinis et al. (2009) and Loh et al. (2010) into a coherent analysis including HOD modelling, and improve the statistics due to the larger survey area included in the newest data release.

Throughout this analysis, we assume a flat ΛCDM cosmology with $\Omega_{\rm m} = 0.3$ and $\sigma_8 = 0.8$. Unless specified otherwise, all distances are coming and quoted in Mpc h^{-1}, and all absolute magnitude are given in $h = 1$ units.

2 DATA

2.1 Sloan Digital Sky Survey

The SDSS (York et al. 2000) mapped most of the high-latitude sky in the northern Galactic cap using a dedicated wide-field 2.5-m telescope at Apache Point Observatory (Gunn et al. 2006) with the SDSS camera (Gunn et al. 1998). The raw imaging data were processed by a series of pipelines performing photometric calibration (Hogg et al. 2001; Ivezić et al. 2004; Tucker et al. 2006), photometric reduction (Lupton et al. 2001) and astrometric calibration (Pier et al. 2003). DR7 (Abazajian et al. 2009) of the spectroscopic sample provides ($u'g'r'i'z'$)-photometry (Fukugita et al. 1996; Smith et al. 2002) and spectra for nearly 900 000 galaxies with $m_r < 17.77$ over 8000 deg^2. These galaxies were selected from the photometric survey for spectroscopic follow-up using specific algorithms for the main galaxy sample (Strauss et al. 2002) and luminous red galaxies (Eisenstein et al. 2001). The main spectroscopic galaxy sample is nearly complete to $r < 17.77$ and has a median redshift of $z \sim 0.1$. Based on these observations, the New York University Value Added Galaxy Catalogue (VAGC; Blanton et al. 2005) contains galaxy samples which have been constructed for large-scale structure studies: all magnitudes are recalibrated (Padmanabhan et al. 2008) and K-corrected (Blanton et al. 2003a), and the radial selection function and angular completeness are carefully determined from the data. We restrict this sample to $m_r < 17.6$ to ensure uniform completeness of faint galaxies across the survey area.

Due to fibre placement in the SDSS spectrograph (Blanton et al. 2003b), galaxies closer than 55 arcsec cannot be observed on the same spectroscopic plate, and hence no redshifts have been measured for about 7 per cent of all targeted galaxies. The lack of observed close galaxy pairs affects the measured correlation functions on small scales. While it is possible to correct for fibre collisions down to 0.01 Mpc h^{-1} (Li et al. 2006), the number density of green valley galaxies is too small to obtain correlation function measurements at such small separations and we simply assign galaxies with missing spectra the redshift of its nearest neighbour. This method has been shown to work well for projected correlation functions above the scale corresponding to 55 arcsec (Zehavi et al. 2005). For the most distant galaxies in our sample, the fibre collision scale is 0.07 comoving Mpc h^{-1} and we measure correlation functions only on perpendicular scales $r_{\rm p} \geq 0.1$ Mpc h^{-1}.

Spectral line measurements and mass estimates for these galaxies are taken from the MPA-JHU catalogue.[2] We use the former to classify the (NUV $-$ r) selected transitional galaxies with emission line diagrams and to compare (NUV $-$ r) colour selection with spectroscopic separation of active and quenched galaxies based on $D_{\rm n}4000$ (Fig. 8). Note that these quantities are estimated from a fibre size of 3 arcsec, and due to low redshift of our galaxy sample these measurements may not be representative of the luminosity averaged properties of a galaxy but rather be dominated by central (bulge dominated) regions.

[1] We note that the increase in effective volume is limited to those regions of the SDSS footprint that are closer to the combined *GALEX*–SDSS footprint than the largest scales probed by the CCF. Due to the patchy geometry of the *GALEX* footprint, these regions cover nearly the entire SDSS footprint.

[2] http://www.strw.leidenuniv.nl/ jarle/SDSS/

2.2 *GALEX*

Near-ultraviolet (NUV) photometry for this project is taken from the *GALEX* Medium Imaging Survey (MIS) Source Catalogue (GMSC; Seibert et al., in preparation) derived from the *GALEX* GR6 data release, which provides unique measurements of point and extended sources up to 1-arcmin diameter in the *GALEX* bands (Seibert et al., in preparation). The NUV source catalogue covers 4827 deg^2 at $\lambda_{\rm eff} = 2316$ Å with a resolution of 5.3 arcsec and reaching a depth ≈23 mag.

GALEX has a circular field of view of 1.°2 which is sampled at 1.5 arcsec. Each field targets a pre-defined position on the sky, resulting in a hexagonal tiling of the survey. These angular selection parameters are contained in exposure time, coverage and flag maps in HEALPIX (Górski et al. 2005) format accompanying the GMSC, which we use to define the combined footprint and select our galaxy sample as detailed in Section 2.3.

2.3 SDSS–MIS cross-match

In order to match the VAGC with NUV detections, we first construct the combined footprint of these two surveys. This is done by converting the VAGC angular selection function, which is given in terms of MANGLE polygons (Hamilton & Tegmark 2004), into the pixelized HEALPIX format (Swanson et al. 2008). Then we multiply the angular selection functions of the VAGC and MIS in each pixel (at resolution $N_{\rm side} = 2048$) and restrict the overlap region to pixels where the angular completeness fraction of both surveys is larger than 0.7. This results in a combined survey with an effective area of 2708 deg^2. Furthermore, we require tiles to have NUV exposure times $t > 1000$ s, which reduces the combined effective area to 1945 deg^2. This final overlap region is shown in black in Fig. 1.

We cross-match all galaxies in the VAGC within this overlap area with NUV detections using a search radius of 4 arcsec. In order to construct a complete statistical sample, we then restrict the cross-match with various cuts summarized in Table 1. Due to deblending and centring issues for nearby or very bright objects, the NUV- and *r*-band photometry pipelines may report positions for these objects that are farther separated than the matching radius, leading to spurious non-detections. Furthermore, the astrometric and photometric precision of the *GALEX* detections declines towards the edges of each tile, and near light echoes and other imaging artefacts and we exclude this regions as detailed in Table 1. The colour–apparent magnitude distribution and completeness of the final cross-match sample are shown in Fig. 2. For apparently bright galaxies ($m_r \lesssim 16$) the blue sequence [around (NUV − *r*) ≈ 2–3] and the red sequence [around (NUV − *r*) ≈ 5–6] are clearly visible. No galaxies are found with (NUV − *r*) $\gtrsim$ 6.5, though these should well be within the *GALEX* detection limit (indicated by the inclined line) at these magnitudes if they existed. For these bright galaxies far from the NUV detection limit the cross-match completeness is around 90 per cent, it decreases for fainter objects as the NUV detection limit moves into the colour–magnitude space occupied by red galaxies. In order to retain a nearly complete sample of green valley galaxies, we cut the cross-match sample at $m_r < 17.1$. The resulting cross-match catalogue has a completeness of 76 per cent, i.e. 76 per cent of galaxies in the VAGC catalogue, which meet the magnitude and redshift criteria described above are at a position with *GALEX* coverage as detailed in Table 1, have a $m_{\rm NUV} < 23.0$ *GALEX* detection.

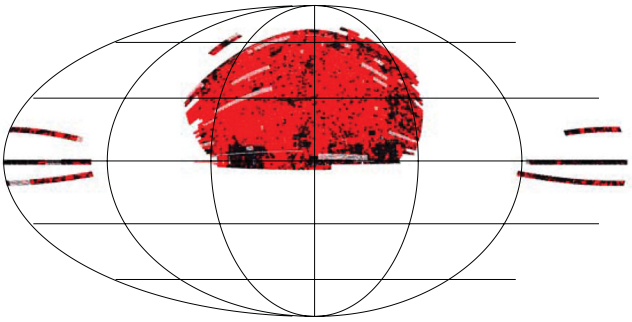

Figure 1. Combined SDSS + *GALEX* MIS footprint. The area covered by the VAGC at an angular completeness $f_{\rm comp} > 0.7$ is shown in red, the final overlap area of 1945 deg^2 between VAGC and MIS, as detailed in Section 2.3, is shown in black.

Table 1. Cross-match sample definition.[a]

Parameter	Limits
r-band magnitude	$14.1 < r < 17.1$
Redshift	$0.02 < z < 0.2$
GALEX field radius	$fov\_radius < 0.^\circ55$
GALEX exposure time	$t > 1000$ s
NUV flag	$nuv\_artifact \leq 1$
NUV magnitude	$16.0 < \mathrm{NUV} < 23.0$
SDSS/NUV angular completeness	$f_{\rm comp} > 0.7$

[a]The parent catalogue is the NYU VAGC `dr72bright`.

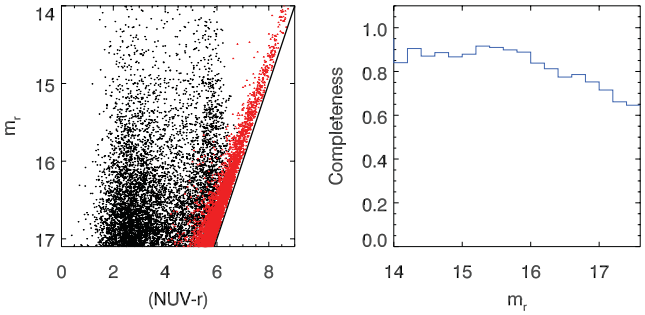

Figure 2. Completeness of the cross-match sample. Left: apparent magnitude–(NUV − *r*) colour diagram. Black dots show a random subset of VAGC galaxies with NUV cross-match. Red dots indicate VAGC galaxies without NUV detections, which have been placed at the detection limit NUV = 23 and corrected for position-dependent galactic extinction. Right: completeness of the NUV cross-match as a function of apparent magnitude.

Finally, we use KCORRECT V4.2 (Blanton & Roweis 2007) to calculate absolute $\mathrm{NUV}_{0.1}$ magnitudes of the cross-match galaxies *k*-corrected to $z = 0.1$. As the redshift evolution in the NUV is not very well constrained, we do not attempt to apply evolution corrections to the NUV nor optical magnitudes. Similarly, we do not attempt to correct the (NUV − *r*) colours for intrinsic extinction. To isolate transitional galaxies and avoid identifying dusty (edge-on) spiral galaxies as green valley objects, we only consider objects with *r*-band isophotal axis ratio $b/a > 0.5$.

3 SAMPLE DEFINITION

In order to work with well-defined galaxy populations, we construct a number of volume-limited samples. As the properties of green valley galaxies may vary with luminosity, we define samples

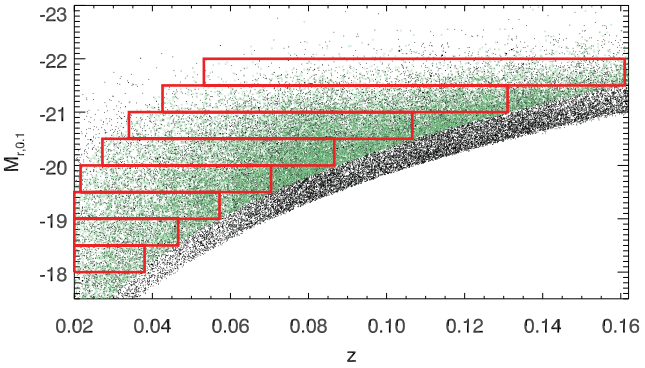

Figure 3. Volume-limited colour selected galaxy samples. Black dots show a random subsample of VAGC galaxies with $m_r < 17.6$, subsampled by a factor 10. Green symbols indicate green valley galaxies identified based on their (NUV $-$ r) colour, which are restricted to $14.1 < m_r < 17.1$ to ensure (near) completeness of the cross-matched sample. Red boxes indicate the location of volume-limited colour selected galaxy samples in luminosity–redshift space.

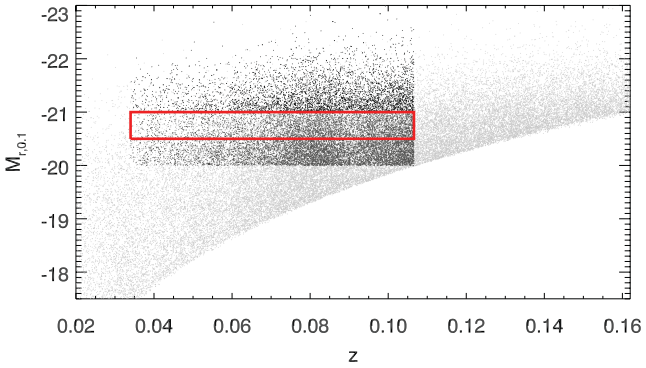

Figure 4. Definition of volume-limited SDSS galaxy reference samples. Light grey dots show a random subsample of VAGC galaxies with $m_r < 17.6$, subsampled by a factor of 10. The red box indicates the location of $[-20.5, -21]$ magnitude range volume-limited colour selected galaxy samples. The dark grey points show the extend of the volume-limited 'faint' SDSS galaxy reference sample associated with this colour selected galaxy sample, the black dots illustrate the associated 'bright' luminosity threshold reference sample. The definitions are analogous for other magnitude ranges, hence we show only one example to improve clarity of the plot.

Table 2. Volume-limited galaxy samples.

Green valley sample				SDSS samples	
M_r	$\langle z \rangle$	N_G	$\bar{n}_G$	N_f	N_b
$[-18, -18.5]$	0.031	285	1.09	15 714	22 177
$[-18.5, -19]$	0.036	595	1.19	24 725	28 488
$[-19, -19.5]$	0.044	869	0.92	38 537	33 041
$[-19.5, -20]$	0.055	1191	0.67	62 193	37 310
$[-20, -20.5]$	0.068	1746	0.54	95 204	36 561
$[-20.5, -21]$	0.083	2028	0.35	109 490	23 586
$[-21, -21.5]$	0.102	1383	0.13	112 647	12 073
$[-21.5, -22]$	0.128	775	0.04	87 676	4458

Note. The first two columns give the magnitude range $[M_{r,\,\mathrm{min}}, M_{r,\,\mathrm{max}}]$ and mean redshift of the green valley galaxy samples illustrated in Fig. 3. N_G is the number of green valley galaxies in this sample, and $\bar{n}_G$ their mean comoving density per 10^{-3} $(\mathrm{Mpc}\,h^{-1})^3$. N_f and N_b are the number of SDSS galaxies in the faint and bright sample in the same volume; the bright sample consists of galaxies in the same volume that are brighter than $M_{r,\,\mathrm{max}}$, and the faint sample contains galaxies in the magnitude range $[M_{r,\,\mathrm{min}} + 0.5, M_{r,\,\mathrm{max}}]$.

of width 0.5 in absolute magnitude, and find the redshift range over which all galaxies in this sample have apparent magnitudes $14.1 < m_r < 17.1$ (the magnitude range of the cross-matched catalogue), cf. Fig. 3. The VAGC has less stringent apparent magnitude requirements ($10 < m_r < 17.6$), and we define two samples of SDSS galaxies occupying the same volume as each luminosity bin sample of NUV detected objects, which are used for the cross-correlation analysis. These samples are described in detail in Table 2. Specifically, for the luminosity bin $[M_{r,\,\mathrm{min}}, M_{r,\,\mathrm{max}}]$ we define the 'bright' sample of SDSS galaxies to contain all galaxies in the same redshift range brighter than $M_{r,\,\mathrm{max}}$, and the 'faint' sample to consist of the volume-limited sample $[M_{r,\,\mathrm{min}} + 0.5, M_{r,\,\mathrm{max}}]$. The definition of these samples luminosity redshift space is illustrated in Fig. 4. We refer to the union of these two samples, which is a luminosity threshold sample with threshold $M_{r,\,\mathrm{min}} + 0.5$, as the SDSS 'all' sample.

3.1 Finding the green valley

We define the location of the green valley in (NUV $-$ r) colour–magnitude space by fitting blue and red sequences to the colour distribution of each volume-limited sample. We include galaxies without NUV detections, which otherwise meet all cross-match criteria and are optically red $[(g - r) > 0.8]$, by placing them at the NUV detection threshold, correcting for position-dependent galactic extinction and assigning the mean k-correction of cross-matched galaxies which are within $\Delta(\mathrm{NUV} - r) = \pm 0.1$ mag, $\Delta M_r = \pm 0.1$ mag and $\Delta z = \pm 0.02$ of the unmatched galaxy. We the find the centre and scatter of the colour sequences by fitting each sequence with a Gaussian. Initially, we cut the distribution at (NUV $-$ r) $= 4.2$ and fit a Gaussian to each side. We then iteratively adjust the fitting range to include the galaxies within 1σ of the peak location on the ridge towards the green valley. The fit distributions are shown in Fig. 5, and the best-fitting parameters for each sample are shown in Fig. 6 along with fits to the blue and red sequences obtained by Wyder et al. (2007), which are based on a different fitting scheme and one continuous galaxy sample weighted by the v_{max} method instead of using disjunct volume-limited samples. As we include NUV non-detections, which are unaccounted for by Wyder et al. (2007), our red sequence is slightly redder for faint galaxies, but otherwise these results agree very well.

The black error bars in Fig. 5 illustrate the mean photometric uncertainty in the (NUV $-$ r) colour of blue/red galaxies, suggesting that asymmetric scatter into the green valley due to photometric uncertainties is small compared to the intrinsic scatter of the red sequence.

3.2 Sample properties

In order to facilitate the comparison with other studies of transitional galaxies based on optical criteria, we characterize the (NUV $-$ r) selected galaxies in other parameter spaces.

Figs 7 and 8 show the distribution of (NUV $-$ r) selected galaxies in ($g - r$) colour space and as a function of the Balmer break index D_n4000. Here the red sample again includes NUV non-detections as described in Section 3.1. The vertical lines indicate the transition between blue/red and star-forming/quenched galaxies based on ($g -$ r) and D_n4000, respectively. Most faint (NUV $-$ r) selected green valley galaxies are optically blue and would be classified as star forming by both of these criteria. On the other hand, a large fraction

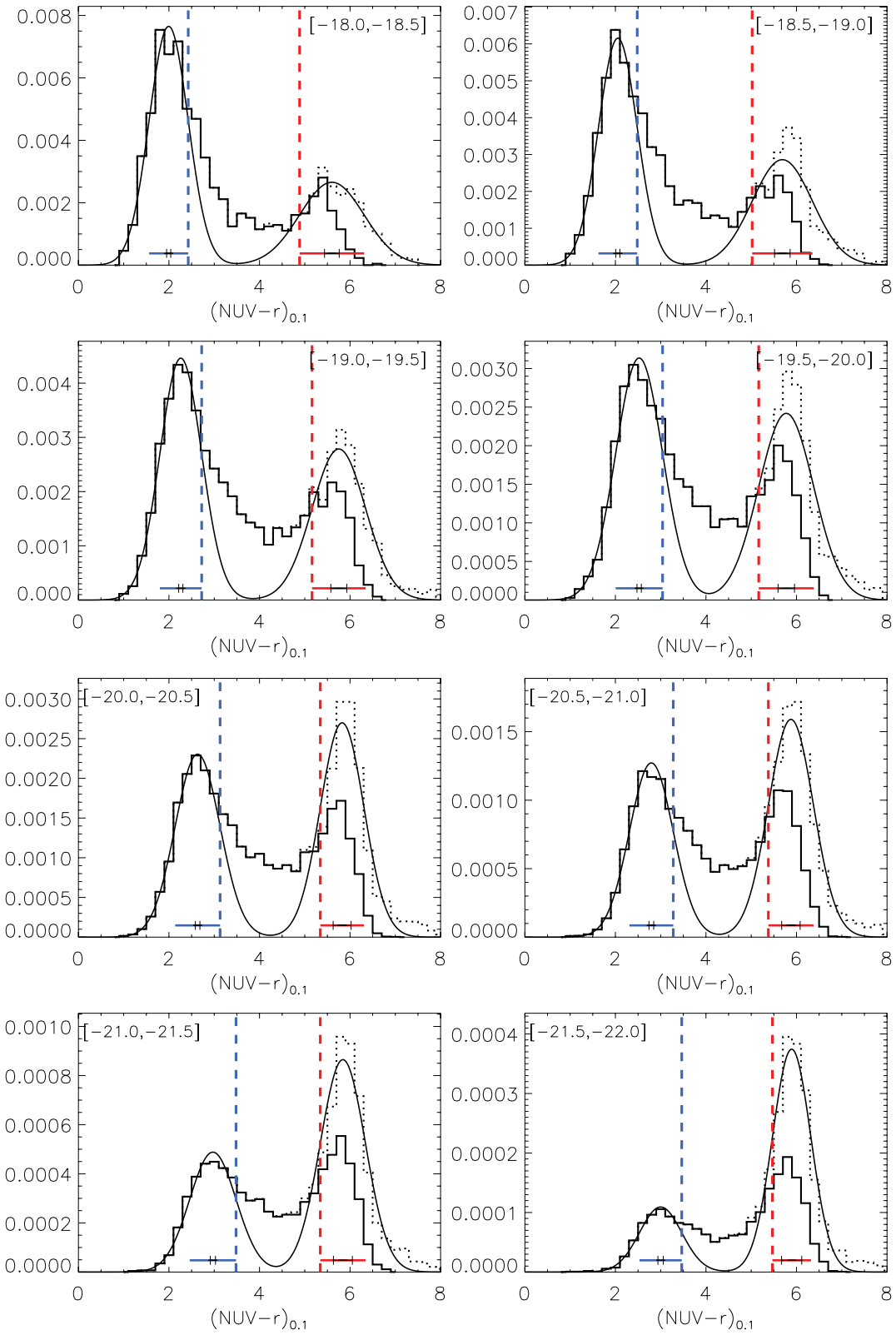

Figure 5. Comoving density of the volume-limited galaxy samples as a function of (NUV − r) colour. Solid histograms show all NUV detected galaxies. The dotted histograms include NUV non-detections, which otherwise meet all cross-match criteria and are optically red ($(g - r) > 0.8$), placed at the NUV detection threshold, corrected for position-dependent galactic extinction and assigned the mean k-correction of cross-matched galaxies which are within $\Delta(\mathrm{NUV} - r) = \pm 0.1$ mag, $\Delta M_r = \pm 0.1$ mag and $\Delta z = \pm 0.02$ of the unmatched galaxy. The solid line shows the double Gaussian fit to the blue side of the blue sequence and the red side of the red sequence, as described in Section 3.1, and the vertical blue and red lines show the 1σ ridge of the colour sequences derived from these fits. The coloured error bars also indicate the 1σ scatter of the colour sequences centred on their respective peak. The black error bars illustrate the mean photometric uncertainty in the (NUV − r) colour of blue/red galaxies.

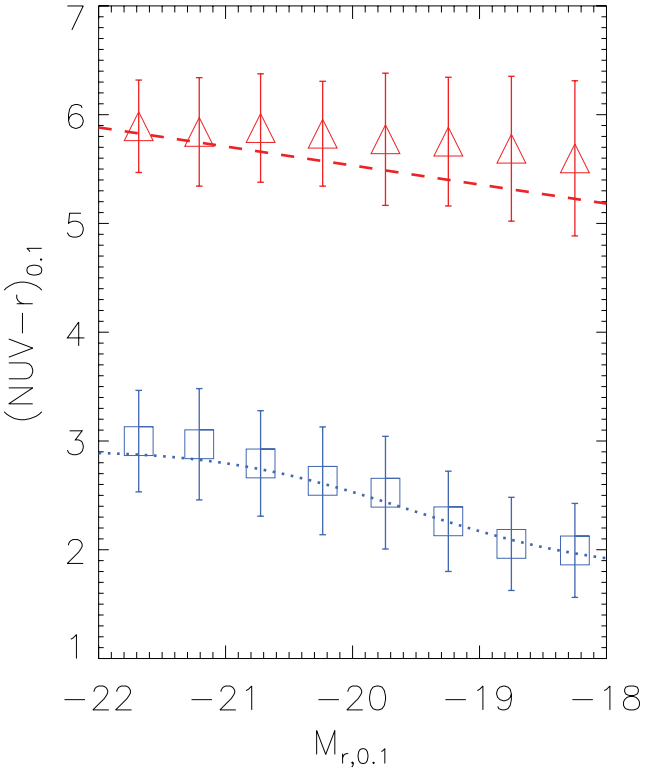

Figure 6. Defining the green valley: symbols and error bars show the location and scatter of the blue and red sequences from the fits in Fig. 5. Lines show the best-fitting sequences from Wyder et al. (2007) transformed to our magnitude units.

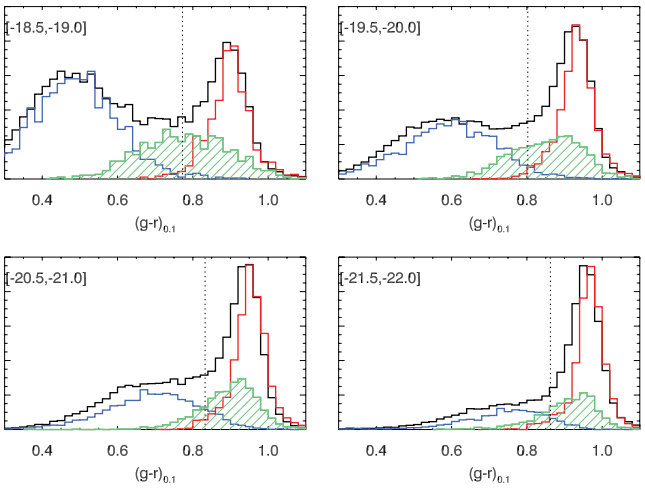

Figure 7. Coloured histograms show the distribution of (NUV − r) selected blue/green/red galaxies in (g − r) space. The black histogram shows the distribution of all SDSS galaxies in the volume-limited sample, but not restricted to the combined footprint. The vertical line shows the colour cut separating blue and red galaxies from Zehavi et al. (2011).

of luminous, (NUV − r) selected transitional galaxies would be classified as red/quenched by both of these criteria.

Furthermore, Fig. 9 shows the distribution of stellar masses as a function of (NUV − r) colour. The stellar masses are taken from the MPA-JHU catalogue and are based on Kauffmann et al. (2003b). At fixed luminosity, green valley galaxies and red sequence galaxies have similar stellar masses.

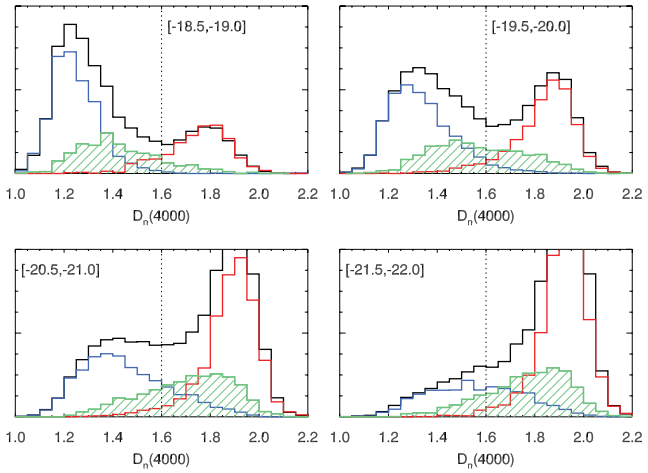

Figure 8. Same as Fig. 7 but for D_n4000. The vertical line shows the separation between quenched ($D_n4000 > 1.6$ and star-forming galaxies used in Tinker, Wetzel & Conroy (2011).

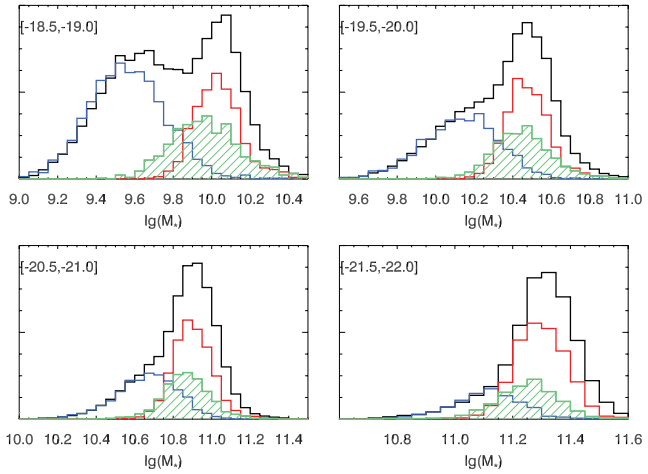

Figure 9. Same as Fig. 7 but for stellar mass.

We illustrate the distribution of green valley galaxy spectra for different luminosity bins in Fig. 10. The thick line shows the mean spectrum obtained from stacking all green valley galaxies (with r-band isophotal axis ratio larger than 0.5) within $\Delta z = 0.02$ of the mean redshift of each luminosity bin. The individual spectra are normalized to the median flux in the 410–500 nm range, giving each galaxy equal weight. The thin grey lines show smoothed individual spectra of 25 galaxies randomly chosen from those used in the stacking process. While we use the spectra mask to exclude pixels flagged by the SDSS spectra reduction pipeline, these spectra contain residual atmospheric [O I] and OH. Note that the fibre diameter of 3 arcsec roughly corresponds to 1.5 kpc h^{-1} at $z = 0.036$, to 3 kpc h^{-1} at $z = 0.083$ and to 4.8 kpc h^{-1} at $z = 0.13$. The stacked spectra show that, on average, green valley galaxies have red bulges and some amount of active galactic nucleus (AGN) activity. All spectra show Hα, or a combination of Hα and [N II], emission, which we classify further using emission line diagnostics in Table 3. For green valley galaxies with emission line measurements with signal-to-noise ratio (S/N) > 3 the AGN fraction is substantial, especially among the more luminous ones. Note that we use emission lines from the MPA-JHU with rescaled flux errors. However, in particular, for the less luminous samples at lower

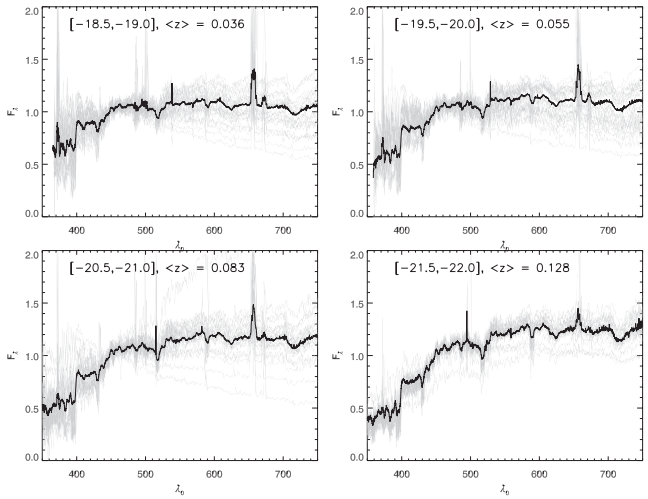

Figure 10. Stacked spectra of (NUV − r) selected green valley galaxies for different luminosity bins as a function of rest-frame wavelength. The thin grey lines show 25 randomly chosen individual spectra, boxcar smoothed over 10 pixel to enhance readability.

Table 3. Classification of green valley galaxies.

M_r	$f_{\mathrm{H}\alpha}$ [a]	f_{SF} [b]	f_{comp} [b]	f_{AGN} [b]	$f_{\mathrm{low\,S/N}}$ [c]
[−18, −18.5]	0.94	0.56	0.04	0.04	0.34
[−18.5, −19]	0.92	0.47	0.14	0.06	0.34
[−19, −19.5]	0.92	0.31	0.22	0.09	0.38
[−19.5, −20]	0.90	0.28	0.25	0.10	0.38
[−20, −20.5]	0.84	0.16	0.20	0.11	0.55
[−20.5, −21]	0.75	0.12	0.16	0.10	0.66
[−21, −21.5]	0.69	0.09	0.11	0.09	0.72
[−21.5, −22]	0.60	0.06	0.09	0.08	0.81

[a]Fraction of green valley galaxies with Hα emission detected at S/N > 3.
[b]Fraction of green valley galaxies classified as star forming (f_{SF}), composite (f_{comp}) or AGN (f_{AGN}) based on the [N II]/Hα versus [O III]/Hβ (Baldwin, Phillips & Terlevich 1981) emission line diagram, using the Kewley et al. (2001) extreme starburst classification line and the Kauffmann et al. (2003a) pure star formation line.
[c] Fraction of galaxies with low S/N (<3) in at least one of these emission lines, not included in the emission line classification.

redshifts, there is considerable spread among objects, limiting the informative value of the stacked spectra.

4 CLUSTERING ANALYSIS

4.1 Projected correlation functions

To separate spatial clustering from redshift space distortions, we first measure the correlation functions in radial direction π and perpendicular direction r_{p} and then project out redshift space distortions. Specifically, we measure the (cross-)correlation function of galaxy samples $D_{X,Y}$ using the Landy & Szalay (1993) estimator and its generalization for CCFs (Szapudi & Szalay 1998):

$$\xi_{XY}(r_{\mathrm{p}},\pi)=\left[\frac{D_XD_Y-D_XR_Y-D_YR_X+R_XR_Y}{R_XR_Y}\right](r_{\mathrm{p}},\pi)\,, \qquad (1)$$

on a 2D grid. Here, $R_{X,Y}$ are associated random catalogues, DD (r_{p},π), DR (r_{p},π) and RR (r_{p},π) are the (normalized) number of data–data, data–random and random–random pairs at separation (r_{p},π). We adopt linear binning in the radial component, logarithmic bins in perpendicular distance and measure the projected (cross-)correlation function as

$$w_{XY}(r_{\mathrm{p}})=2\int_0^{\pi_{\mathrm{max}}}\mathrm{d}\,\pi\;\xi_{XY}(r_{\mathrm{p}},\pi)\,, \qquad (2)$$

with $\pi_{\mathrm{max}}=50\,\mathrm{Mpc}\,h^{-1}$.

4.2 Measurements

We generate random catalogues with the SDSS angular selection function and the angular selection function of the *GALEX*–SDSS cross-match catalogue. As we have constructed volume-limited galaxy samples, and their colour selected subsamples, with narrow redshift ranges allowing us to ignore redshift evolution effects, the random catalogues have uniform comoving density and do not need to account for the radial selection function. The random catalogues are oversampled compared to the galaxy catalogues by a factor 25 for SDSS samples, and by a factor 100 for the sparser (NUV − r) selected samples. Increasing the oversampling rate by a factor of 2 has no significant impact, indicating that the correlation function estimates have converged.

Fig. 11 demonstrates that we have characterized the combined survey geometry sufficiently well to measure correlation functions in this patchy survey geometry. Here we show the correlation function between a galaxy sample in the full SDSS footprint in the magnitude bin [−19.5, −20] and blue colour [$(g-r)<0.8$] with different subsets of itself. The dashed line shows its autocorrelation function. Next we consider the cross-correlation between this sample and its restriction to the footprint of the SDSS+*GALEX* combined catalogue, which is shown by the dotted line. Compared to the full autocorrelation function, this CCF may be affected by boundary effects associated with the correlation function estimator or finite volume effects, as we have reduced the volume probed by

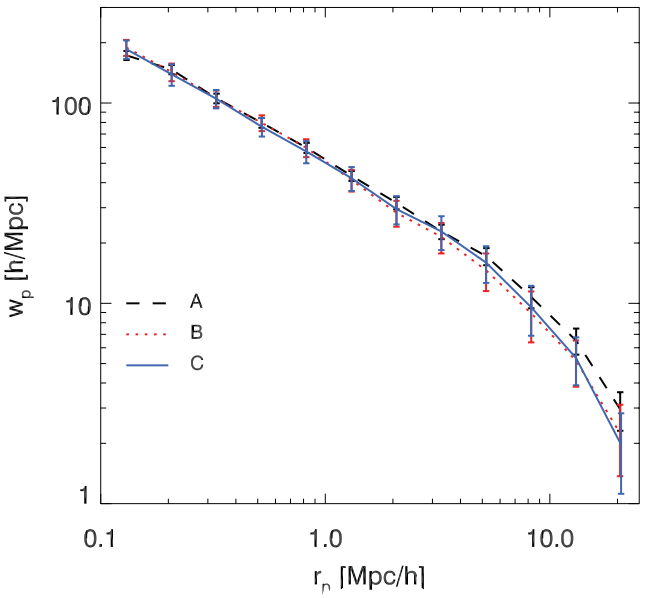

Figure 11. Test of survey geometry effects on measured correlation functions. Different lines show the projected CCF between galaxies in the full SDSS footprint in the magnitude bin [−19.5, −20] and with $(g-r)<0.8$ (A) with the same sample, (B) with the sample restricted to the combined survey area, (C) with *GALEX*-detected galaxies in the same magnitude and colour bin.

of one copy of the galaxy catalogue by a factor of 4. Note that in this case the angular selection function in the combined survey area is still given by the SDSS angular selection function. Next we further restrict one copy of the galaxy catalogue to galaxies with NUV detections, shown by the solid line. As the galaxy sample consists only of blue galaxies, these should all have NUV detections, and any significant differences between the dotted and solid line would indicate a mischaracterization of the combined angular selection function. One copy of the galaxy catalogue stays the same throughout the process, so that we measure the cross-correlation between samples with different footprints, which leads to better statistics and smaller finite volume effects than restricting the SDSS data to the combined footprint region as well.

As described in detail in Zehavi et al. (2011), the clustering of the faintest SDSS luminosity threshold samples is subject to substantial sample variance effects due to the small volume probed by these low-redshift samples. As we are interested in a sparse subpopulation of these samples and are furthermore restricted to one-fourth of the SDSS footprint area, these sampling effects are even more severe in our analysis. After reproducing the subvolume tests of Zehavi et al. (2011), we find that the magnitude bin $[-19.5, -20]$ is the smallest sample for which we can obtain robust correlation function measurements. Examples of measured autocorrelation function and CCF for SDSS galaxy samples and green valley galaxies are shown in Fig. 12. For comparison, we also show measurements the green valley galaxy autocorrelation function, for which we used random catalogues with an oversampling factor of 1000.

We estimate the covariance of our correlation function measurements using bootstrapping with 'oversampling of subvolumes' (Norberg et al. 2009) with an oversampling factor of 3, where number of subvolumes chosen with replacement N_r is equal to three times the number of subvolumes the data set is divided up into, $N_{\rm sub}$. Norberg et al. (2009) find that this method gives robust error estimates that are in agreement with external estimates from mock catalogues. For correlation functions between two SDSS galaxy samples, we divide the SDSS footprint into 150 subsets of equal area. For correlation functions between one SDSS galaxy sample and one sample restricted to the combined footprint area, the division into equal area subsets is not clearly defined, and we choose subsets which contain equal number of random–random pairs at angular separation of 2° in order to evenly sample the CCF on scales of a few Mpc h^{-1}. Due to the smaller effective area of this restricted geometry, we only have 50 such subareas. Examples for both types of covariances are shown in Fig. 13. As noted by Hartlap, Simon & Schneider (2007), estimated covariances are a biased estimate of the inverse covariance with the bias depending on the number of data points, p, and the number of independent data sets, n. If the mean is estimated from the data, an unbiased estimate of the inverse covariance is given by

$$\hat{\boldsymbol{C}}^{-1}_{\rm unbiased} = \frac{n-p-2}{n-1}\hat{\boldsymbol{C}}^{-1} \approx \left(1-\frac{p}{n}\right)\hat{\boldsymbol{C}}^{-1}. \quad (3)$$

As bootstrap realizations are not independent, we cannot apply equation (3) directly with $n = N_r$. Instead, we assume

$$\hat{\boldsymbol{C}}^{-1}_{\rm unbiased} \approx \left(1-m\frac{p}{N_r}\right)\hat{\boldsymbol{C}}^{-1}, \quad (4)$$

and follow the calibration method described in Eifler, Kilbinger & Schneider (2008). We measure $\mathrm{tr}(\hat{\boldsymbol{C}}^{-1})$ repeatedly varying N_r with constant binning and oversampling rate, and determine m as the slope of $1/\mathrm{tr}(\hat{\boldsymbol{C}}^{-1})$ with in p/N_r. Specifically, we varied N_r using $N_{\rm sub} = (120, 135, 150, 165, 180)$ for the SDSS footprint, and $N_{\rm sub} = (40, 45, 50, 55, 60)$ for the *GALEX*–SDSS footprint.

We were unable to obtain stable, invertible covariances for the most luminous green valley galaxy sample. Hence we restrict our analysis of this sample to large scales (Section 4.3) where it was possible to measure converged and invertible covariances.

4.3 Results: large-scale bias

Based on the correlation function measurements described in the previous section, we can measure the large-scale galaxy bias by fitting the projected correlation functions with theoretical matter correlation functions times a linear bias factor. Specifically, we fit measured correlation functions over the range 3–25 Mpc h^{-1} to the theoretical predictions for the projected matter correlation function, including the full data covariance. Fig. 14 shows the resulting luminosity bias relation. The top two plots are for binned and threshold samples of SDSS galaxies, and the lines are fits from the analysis of galaxy clustering in SDSS DR7 by Zehavi et al. (2011). Overall, we find good agreement with their results. The $M_r < -20$ galaxy threshold sample and its subsamples deviate from the best-fitting bias relation. As detailed in Table 2, these samples are centred around the redshift of the Sloan Great Wall, which leads to excess clustering in this and neighbouring samples.[3] This effect is enhanced in the lower plots, which show bias as a function of (NUV $- r$) colour and luminosity or mean stellar mass. Here the clustering of red galaxies is strongly enhanced in the Sloan Great Wall.

5 HALO OCCUPATION DISTRIBUTION MODELLING

At the level of individual haloes, a HOD model (e.g. Berlind & Weinberg 2002) describes the relation between galaxies and halo mass in terms of the probability $P(N, M_{\rm h})$ that a halo of given mass $M_{\rm h}$ contains N galaxies. To describe the two-point clustering of galaxies, we need models for first and second moment of the HOD, $\langle N|M_{\rm h}\rangle$ and $\langle N(N-1)|M_{\rm h}\rangle$. Following Zheng et al. (2005), we separate galaxies into central and satellite galaxies. By definition, a halo contains either zero or one central galaxy, and it may host satellite galaxy only if it contains a central galaxy, which motivates the form:

$$\langle N(M_{\rm h})\rangle = \langle N_{\rm c}|M_{\rm h}\rangle\,(1+\langle N_{\rm s}|M_{\rm h}\rangle)\,, \quad (5)$$

with $\langle N_{\rm c/s}|M_{\rm h}\rangle$ the average number of central/satellite galaxies in a halo of mass $M_{\rm h}$.

5.1 HOD parametrization

While the assumptions in a HOD model describing the properties of dark matter haloes are generally agreed upon (see Section 5.2 for details), the form of the relation between galaxies and haloes (equation 6) is less well constrained and leaves more room for experiments. We motivate the details our implementation next.

5.1.1 SDSS samples

We base our model for SDSS galaxy samples on the HOD parametrization of Zehavi et al. (2011) for luminosity thresholds

[3] This was also noted by Zehavi et al. (2011) who exclude the redshift range of the Sloan Great Wall from their analysis of luminosity bin galaxy samples.

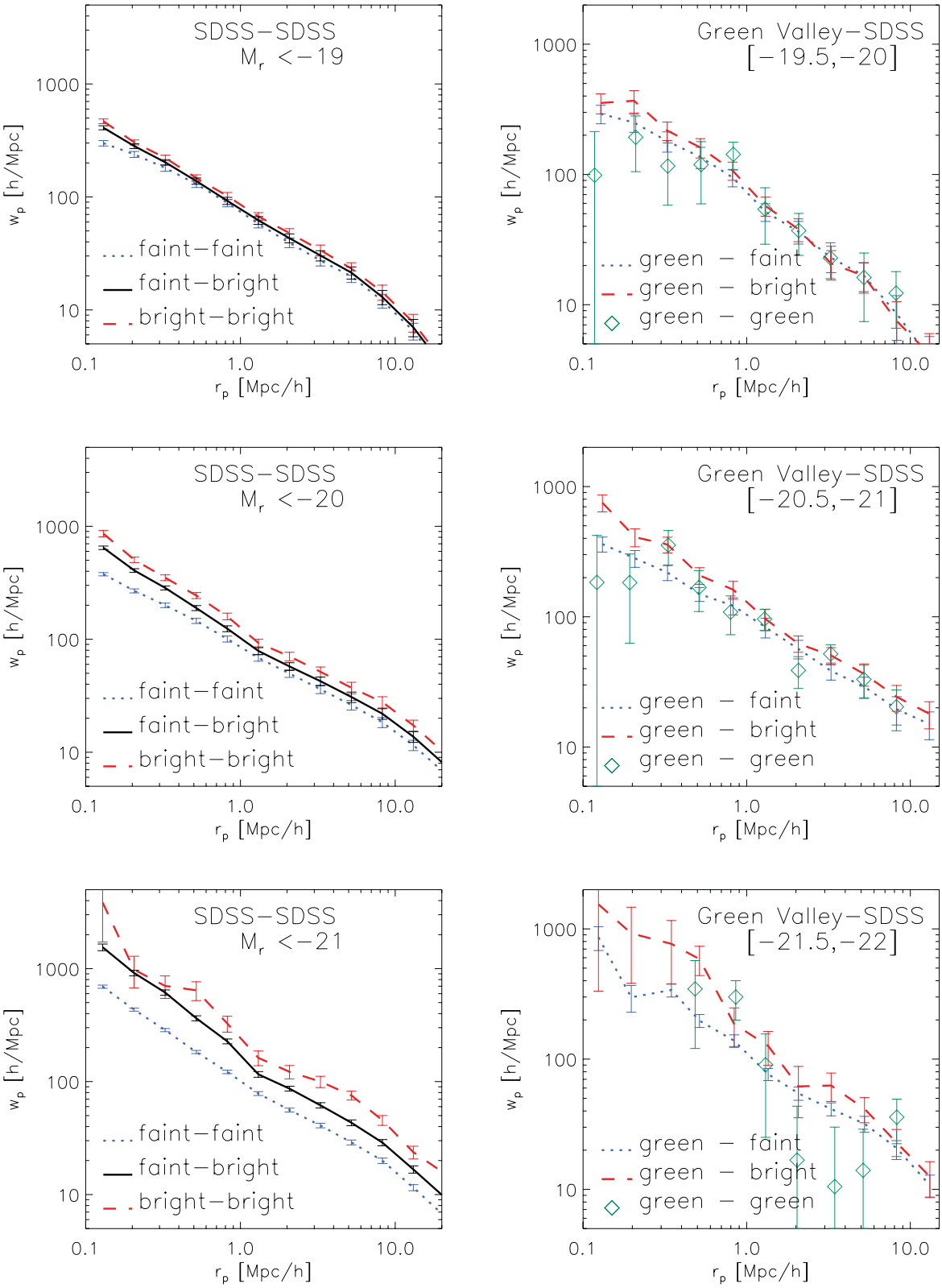

Figure 12. Examples of measured CCFs. For comparison, we also show measurements of the green valley galaxy autocorrelation function.

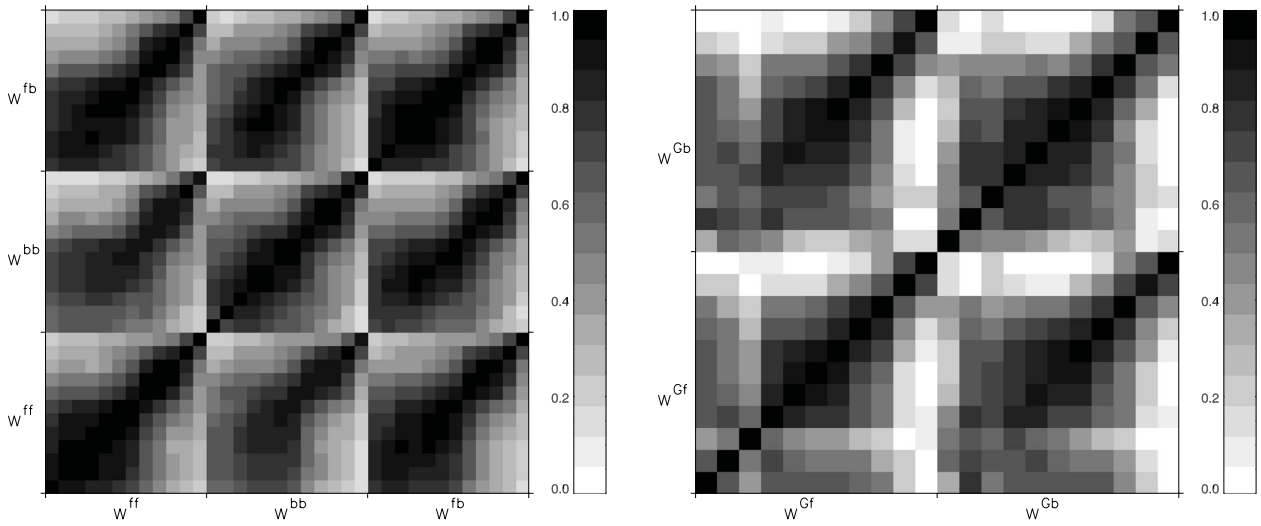

Figure 13. Sample covariances. Left: covariance between the different autocorrelation function and CCFs of the SDSS faint and bright samples associated with the magnitude bin [−19.5, −20]. Right: covariance of the CCF between the [−19.5, −20] green valley sample and the corresponding SDSS faint and bright samples. In each block of these covariances, perpendicular scales increase from left to right and bottom to top.

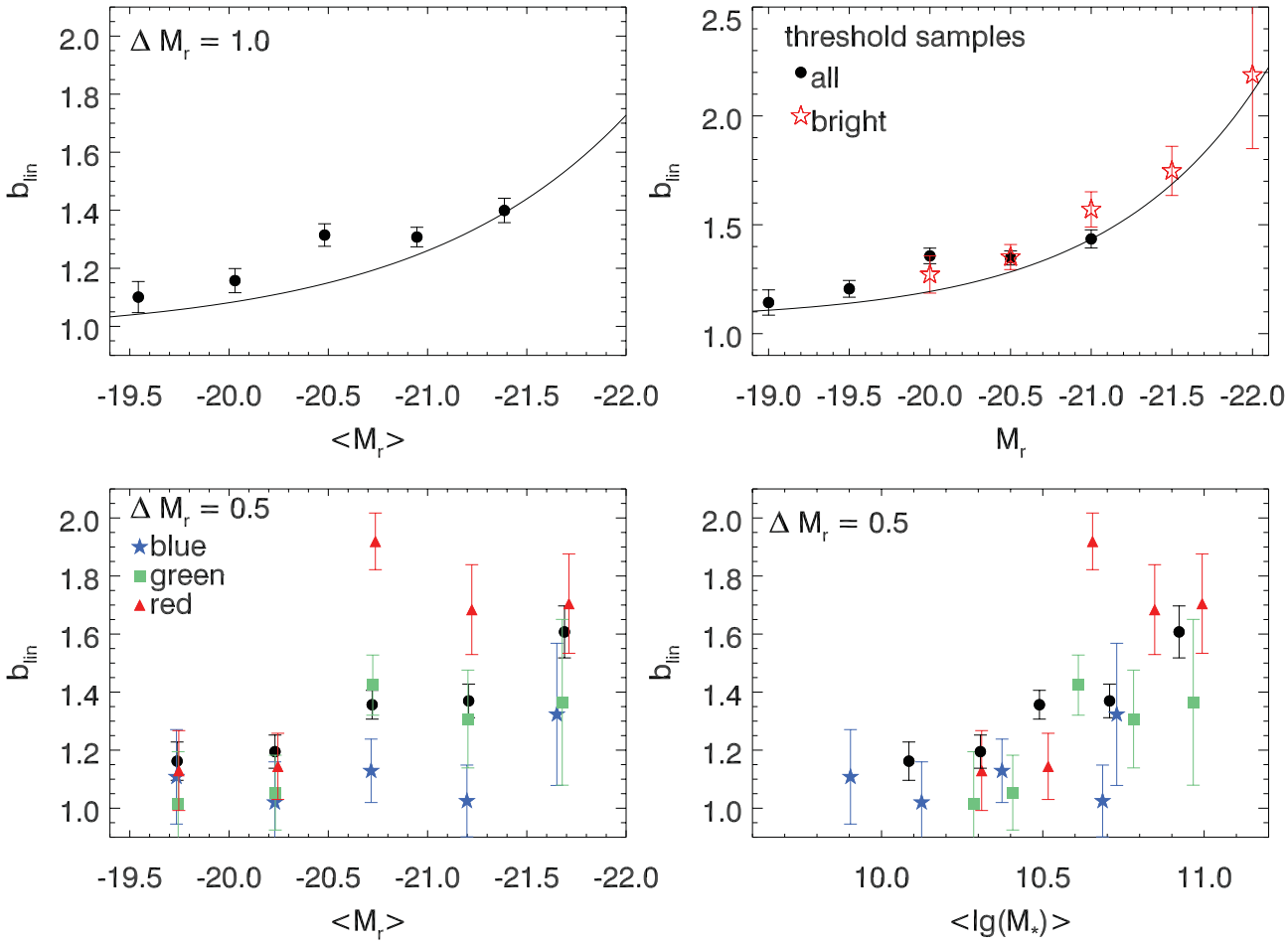

Figure 14. Linear galaxy bias measurements obtained from fits to the large-scale correlation function. Top: linear bias as a function of luminosity for different luminosity bin samples with bin width $\Delta M_r = 1.0$ (left), and threshold samples (right). The lines show best-fitting relations from Zehavi et al. (2011). Bottom: linear bias as a function of (NUV − r) colour and luminosity (left) or stellar mass (right), for galaxy samples with luminosity bin width $\Delta M_r = 0.5$.

samples with absolute r-band magnitude $M_r < M_r^{\rm t}$,

$$\left\langle N(M_{\rm h}|M_r^{\rm t})\right\rangle = \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M_{\rm h}-\log M_{\rm min}^{\rm t}}{\sigma_{\log M}^{\rm t}}\right)\right] \times \left[1+\left(\frac{M_{\rm h}-M_0^{\rm t}}{M_1^{\prime{\rm t}}}\right)^{\alpha_{\rm t}}\right], \quad (6)$$

with model parameters $M_{\rm min}^{\rm t}$, $M_0^{\rm t}$, $M_1^{\prime{\rm t}}$, $\sigma_{\log M}^{\rm t}$ and $\alpha_{\rm t}$. The central galaxy occupation function is a softened step function with transition mass scale $M_{\rm min}^{\rm t}$, which is the halo mass in which the median central galaxy luminosity corresponds to the luminosity threshold, and softening parameter $\sigma_{\log M}^{\rm t}$ which is related to the scatter between galaxy luminosity and halo mass. The normalization of the satellite occupation function, $M_1^{\prime{\rm t}}$, and cut-off scale $M_0^{\rm t}$ are related to M_1, the mass scale at which a halo hosts at least on satellite galaxy $[N_{\rm s}(M_1)=1)]$; finally $\alpha_{\rm t}$ is the high-mass slope of the satellite occupation function. This parametrization was found to reproduce the clustering of SDSS and CFHTLS galaxies (Coupon et al. 2012) well over a large range of luminosity thresholds and redshifts.

The HOD model for a binned galaxy sample with $M_r^{\rm t_2} < M_r < M_r^{\rm t_1}$ is typically calculated from model fits to luminosity threshold samples as

$$\left\langle N\left(M_{\rm h}|M_r^{\rm t_1}, M_r^{\rm t_2}\right)\right\rangle = \left\langle N\left(M_{\rm h}|M_r^{\rm t_1}\right)\right\rangle - \left\langle N\left(M_{\rm h}|M_r^{\rm t_2}\right)\right\rangle . \quad (7)$$

While we note that the results of Zehavi et al. (2011) favour a somewhat steeper slope of the satellite distribution for the most luminous galaxy samples in our analysis, we set $\alpha = 1$ for all SDSS galaxy samples. This is in overall agreement with previous results for the luminosity range of interest, and makes differencing the HOD of neighbouring samples numerically more stable. Hence our model has four free parameters for a luminosity threshold sample, and eight free parameters for a luminosity bin sample. Without further constraints, such a parametrization of luminosity bin samples has too many degrees of freedom for general applications. However, it has the advantage that the HODs of neighbouring luminosity bins are consistent with each other, and we use this parametrization to fit the different correlation functions among our SDSS faint and bright samples, resulting in eight parameters for the SDSS HODs in each volume-limited sample.

Furthermore, we assume the radial distribution of all colour-independent galaxy samples to follow the dark matter distribution. This assumption is supported by the results of Watson et al. (2012) who studied the small-scale clustering of SDSS galaxies. While these authors found an enhanced clustering of luminous galaxies on small scales ($r_{\rm p} < 0.05\,{\rm Mpc}\,h^{-1}$) compared to an Navarro–Frenk–White (NFW) distribution, their galaxy correlation function measurements agree very well with the predicted dark matter clustering over the radial scales and luminosity range of interest for our analysis.

5.1.2 *Luminosity and colour bin samples*

For a (NUV $-$ r) selected galaxy sample (X), which is measured in one narrow 0.5-mag bin per sample volume, we need a more compact description of the HOD and we model the central galaxy term as a clipped Gaussian:

$$\langle N_{\rm c}(M_{\rm h}, X)\rangle = \min\left(\frac{A_X}{\sigma_X}\sqrt{2\pi}\exp\left(\frac{-(\log M_{\rm h}-\log M_{\rm c}^X)^2}{2\sigma_X^2}\right), 1\right), \quad (8)$$

with free parameters A_X, σ_X and $M_{\rm c}^X$. Here the clipping enforces that a halo does not have more than one central galaxy.

The autocorrelation function of colour selected galaxies by definition is only sensitive to galaxy pairs of the same colour. Hence HOD models require assumptions on the relation between the colours of central and satellite galaxies, and in particular need to account for central galaxies which are not part of sample (e.g. Simon et al. 2009; Skibba & Sheth 2009). In contrast, modelling the cross-correlation between a colour selected galaxy sample and the full (colour-independent) galaxy population with the same luminosity threshold does not require such assumptions. This allows us to simply write the condition that a halo has to contain a central galaxy in order to host satellite galaxies in terms of central galaxy occupation function of the full (colour-independent) luminosity threshold sample with luminosity threshold t_X equal to the minimum luminosity of the luminosity bin under consideration:

$$\langle N_{\rm s}(M_{\rm h}, X)\rangle = A_X\frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M_{\rm h}-\log M_{\rm min}^{\rm t_X}}{\sigma_{\log M}^{\rm t_X}}\right)\right] \times \left(\frac{M_{\rm h}}{M_1^X}\right)^{\alpha_X}, \quad (9)$$

which is characterized by two free parameters, M_1^X and α_X.

Note that the correlation function of a binned sample is independent of the normalization parameter A_X, which is determined by the galaxy number density.

Motivated by observations finding red satellite galaxies to be radially more concentrated than blue galaxies (e.g. von der Linden et al. 2010; Guo et al. 2012), we introduce another free parameter f_X which describes the NFW concentration, c_X, of a colour selected galaxy sample relative to that of dark matter:

$$c_X(M_{\rm h}) = f_X c(M_{\rm h}) . \quad (10)$$

5.2 Relation to correlation functions

The halo model prediction for the real-space correlation function takes the form:

$$1+\xi(r) = \left[1+\xi^{\rm 1h}(r)\right] + \left[1+\xi^{\rm 2h}(r)\right], \quad (11)$$

where $(1+\xi^{\rm 1h})$ is proportional to the number of galaxy pairs residing in the same halo (*one-halo term*), and the *two-halo term* $(1+\xi^{\rm 2h})$ is proportional to the number of galaxy pairs occupying different haloes. The model real-space correlation function is related to the projected correlation function as

$$w(r_{\rm p}) = 2\int_0^{\pi_{\rm max}} {\rm d}\pi\, \xi\left(\sqrt{r_{\rm p}^2+\pi^2}\right) . \quad (12)$$

We will now describe the computation of these terms in detail. In order to evaluate these expressions numerically, we define haloes to enclose a spherical overdensity of 200 times the mean background density and assume that their density distribution follows a NWF profile (Navarro, Frenk & White 1997) with the halo mass–concentration relation of Bhattacharya, Habib & Heitmann (2011); furthermore, we use the fitting functions of Tinker et al. (2008) and Tinker et al. (2010) for the halo mass function and halo bias relation. Unless stated otherwise, we assume that the galaxy distribution follows the halo density profile.

5.2.1 One-halo term

We split the computation of the one-halo term into then clustering of central and satellite galaxy $\xi^{1,\mathrm{c-s}}$ and satellite–satellite clustering $\xi^{1,\mathrm{s-s}}$ within the same halo. The central–satellite term is given by

$$1+\xi^{1,\mathrm{c-s}}_{XY}(r) = \frac{1}{\bar{n}_X\bar{n}_Y}\int_{M_{\mathrm{vir}(r)}}^{\infty}\mathrm{d}M_\mathrm{h}\,\frac{\mathrm{d}n}{\mathrm{d}M_\mathrm{h}}(\langle N_\mathrm{c}(M_\mathrm{h},X)N_\mathrm{s}(M_\mathrm{h},Y)\rangle\,\rho_Y(r|M_\mathrm{h}) + \left\langle N_\mathrm{c}(M_\mathrm{h},Y)N_\mathrm{s}(M_\mathrm{h},X)\right\rangle\rho_X(r|M_\mathrm{h}))\,, \quad (13)$$

where $\mathrm{d}n/\mathrm{d}M_\mathrm{h}$ denotes the halo mass function, with $\rho_X(r|M_\mathrm{h})$ the normalized radial distribution of galaxy population X within the halo, and with

$$\bar{n}_X = \int_0^\infty \mathrm{d}M_\mathrm{h}\,\frac{\mathrm{d}n}{\mathrm{d}M_\mathrm{h}}\,\langle N(M_\mathrm{h}|X)\rangle\ . \quad (14)$$

The term $\langle N_\mathrm{c}(M_\mathrm{h},X)N_\mathrm{s}(M_\mathrm{h},Y)\rangle$ in equation (13) is equal to the average number galaxy pairs with a central galaxy from sample X and a satellite galaxy from sample Y in a halo of mass M_h. From the definition of satellite galaxy, this term evaluates to $\left\langle N_\mathrm{c}(M_\mathrm{h},M_r^\mathrm{t})N_\mathrm{s}(M_\mathrm{h},M_r^\mathrm{t})\right\rangle = \left\langle N_\mathrm{s}|M_\mathrm{h},M_r^\mathrm{t})\right\rangle$ for the autocorrelation of a luminosity threshold sample (Zheng et al. 2005). However, when considering binned samples or cross-correlations between different samples, the central galaxy of a halo hosting satellite galaxies from the sample Y need not be from sample X, and we use $\langle N_\mathrm{c}(M_\mathrm{h},X)N_\mathrm{s}(M_\mathrm{h},Y)\rangle = \langle N_\mathrm{c}|M_\mathrm{h},X\rangle\langle N_\mathrm{s}|M_\mathrm{h},Y\rangle$ (Miyaji et al. 2011).

If samples X and Y are disjunct, the satellite–satellite term is given by

$$1+\xi^{1,\mathrm{s-s}}_{XY}(r) = \frac{1}{\bar{n}_X\bar{n}_Y}\int_{M_{\mathrm{vir}(r)}}^{\infty}\mathrm{d}M_\mathrm{h}\frac{\mathrm{d}n}{\mathrm{d}M_\mathrm{h}}\,\langle N_\mathrm{s}(M_\mathrm{h},X)N_\mathrm{s}(M_\mathrm{h},Y)\rangle \times(\rho_X*\rho_Y)(r|M_\mathrm{h})\,, \quad (15)$$

where $(\rho_X*\rho_Y)(r|M_\mathrm{h})$ denotes the convolution of radial galaxy distributions ρ_X and ρ_Y, and where the average number of satellite pairs is given by $\langle N_\mathrm{s}(M_\mathrm{h},X)N_\mathrm{s}(M_\mathrm{h},Y)\rangle = \langle N_\mathrm{s}|M_\mathrm{h},X\rangle\langle N_\mathrm{s}|M_\mathrm{h},Y\rangle$.

To model autocorrelations function, the number of galaxy pairs is modified to

$$1+\xi^{1,\mathrm{s-s}}_{XX}(r) = \frac{2}{\bar{n}_X\bar{n}_X}\int_{M_{\mathrm{vir}(r)}}^{\infty}\mathrm{d}M_\mathrm{h}\frac{\mathrm{d}n}{\mathrm{d}M_\mathrm{h}}\,\frac{\langle N_\mathrm{s}(M_\mathrm{h},X)(N_\mathrm{s}(M_\mathrm{h},X)-1)\rangle}{2} \times\ (\rho_X*\rho_X)(r|M_\mathrm{h})\,. \quad (16)$$

Assuming that satellite galaxies are Poisson distributed, the number of pairs evaluates to $\langle N_\mathrm{s}(N_\mathrm{s}-1)\rangle = \langle N_\mathrm{s}\rangle^2$.

5.2.2 Two-halo term

On scales above $\sim$5Mpc h^{-1}, the clustering of galaxies follows the large-scale clustering of dark matter haloes, and it is modelled as function of the dark matter correlation function ξ^mm,

$$\xi^{2\mathrm{h}}_{XY}(r) \approx b_X b_Y \xi^\mathrm{mm}(r)\,. \quad (17)$$

Here b_X denotes the bias parameter of galaxy sample X, which we calculate as

$$b_X = \frac{1}{\bar{n}_X}\int_0^\infty \mathrm{d}M_\mathrm{h}\,\frac{\mathrm{d}n}{\mathrm{d}M_\mathrm{h}}b_\mathrm{h}(M_\mathrm{h})\,\langle N(M_\mathrm{h}|X)\rangle\ , \quad (18)$$

where b_h is the halo bias parameter.

On intermediate scales, one needs to account for the distribution of galaxies within different haloes and halo exclusion, i.e. the fact that two haloes' contribution to the two-halo term does not overlap. Following the spherical halo exclusion model of Tinker et al. (2005), we restrict the calculation of the two-halo term at separation r to haloes with $R_{\mathrm{vir},1} + R_{\mathrm{vir},2} \leq r$. The effect of the distribution of galaxies within the different haloes on the correlation function is given by the convolution of their respective density profiles. As this requires convolving many different halo profiles, we calculate the two-halo term as calculated in Fourier space:

$$P^{2\mathrm{h}}_{XY}(k,r) = P_m(k)\frac{1}{\bar{n}'_X\bar{n}'_Y(r)} \times\int_{M_\mathrm{min}}^{M_{\mathrm{lim},1}(r)}\mathrm{d}M_1\,\frac{\mathrm{d}n}{\mathrm{d}M_1}\,\langle N|M_1,X\rangle\,b_\mathrm{h}(M_1)\tilde{\rho}_X(k,M_1) \times\int_{M_\mathrm{min}}^{M_{\mathrm{lim},2}(M_1,r)}\mathrm{d}M_2\,\frac{\mathrm{d}n}{\mathrm{d}M_2}\,\langle N|M_2,Y\rangle\,b_\mathrm{h}(M_2)\tilde{\rho}^*_Y(k,M_2)\,, \quad (19)$$

where $M_{\mathrm{lim},1}$ is the maximum halo mass such that $R_\mathrm{vir}(M_{\mathrm{lim},1}) = r - R_\mathrm{vir}(M_\mathrm{min})$ with M_min the minimum halo mass of the HOD, $M_{\mathrm{lim},2}$ is defined by $R_\mathrm{vir}(M_{\mathrm{lim},2}) = r - R_\mathrm{vir}(M_{\mathrm{lim},1})$ and $\tilde{\rho}_X$ denotes the Fourier transform of the normalized galaxy distribution ρ_X. $\bar{n}'_X\bar{n}'_Y(r)$ denotes the number density of galaxy pairs restricted to non-overlapping haloes at separation r:

$$\bar{n}'_X\bar{n}'_Y(r) = \int_{M_\mathrm{min}}^{M_{\mathrm{lim},1}(r)}\mathrm{d}M_1\,\frac{\mathrm{d}n}{\mathrm{d}M_1}\,\langle N|M_1,X\rangle \times\int_{M_\mathrm{min}}^{M_{\mathrm{lim},2}(M_1,r)}\mathrm{d}M_2\,\frac{\mathrm{d}n}{\mathrm{d}M_2}\,\langle N|M_2,Y\rangle\ . \quad (20)$$

The two-halo correlation function is obtained from the power spectrum by

$$\xi^{2\mathrm{h}\prime}_{XY}(r) = \frac{1}{2\pi^2}\int_0^\infty \mathrm{d}k\,k^2\frac{\sin(kr)}{kr}P^{2\mathrm{h}}_{XY}(k,r)\,. \quad (21)$$

As $\xi^{2\mathrm{h}\prime}_{XY}(r)$ has been obtained from a (radius-) restricted sample of galaxy pairs, it is converted to a probability for the whole sample by

$$1+\xi^{2\mathrm{h}}_{XY}(r) = \frac{\bar{n}'_X\bar{n}'_Y(r)}{\bar{n}_X\bar{n}_Y}\left[1+\xi^{2\mathrm{h}\prime}_{XY}(r)\right]. \quad (22)$$

5.3 Analysis

As described in Section 4.2, for each luminosity bin sample of interest we measure the projected autocorrelation function and CCF of the SDSS faint and bright galaxy samples:

$$(\boldsymbol{w}_\mathrm{ff},\boldsymbol{w}_\mathrm{fb},\boldsymbol{w}_\mathrm{bb}) \equiv \boldsymbol{W}_\mathrm{S}\,, \quad (23)$$

where we have introduced the correlation function data vector $\boldsymbol{w} = [w(r_{\mathrm{p},1}), w(r_{\mathrm{p},2}), ..., w(r_{\mathrm{p},N_\mathrm{bin}})]$, and the cross-correlation between (NUV $-$ r) colour selected luminosity bin samples and the two SDSS galaxy samples:

$$(\boldsymbol{w}_{X\mathrm{f}},\boldsymbol{w}_{X\mathrm{b}}) \equiv \boldsymbol{W}_X\,, \quad (24)$$

with $X \in \{\mathrm{blue},\mathrm{green},\mathrm{red}\}$).

Ideally, one would fit all these CCFs simultaneously, however this is not practicable: as the (NUV $-$ r) selected galaxy samples are restricted to *GALEX*+SDSS overlap area, obtaining a joint covariance for the SDSS reference samples and the colour selected sample ($\mathrm{Cov}(\boldsymbol{w}_\mathrm{ff},\boldsymbol{w}_\mathrm{fb},\boldsymbol{w}_\mathrm{bb},\boldsymbol{w}_{X\mathrm{f}},\boldsymbol{w}_{X\mathrm{b}})$) would require restricting the SDSS

Table 4. Best-fitting HOD model parameters for SDSS samples.

$\log M_r$	$\log M^{\rm f}_{\rm min}$	$\sigma^{\rm f}_{\log M}$	$\log M^{\rm f}_0$	$\log M'^{\rm f}_1$	χ^2/d.o.f.
[−19.5, −20.0]	11.55 ± 0.04	0.24 ± 0.11	10.14 ± 0.15	12.80 ± 0.03	1.80
[−20.0, −20.5]	11.64 ± 0.02	0.16 ± 0.08	10.07 ± 0.14	12.92 ± 0.03	3.04
[−20.5, −21.0]	11.98 ± 0.09	0.42 ± 0.13	9.45 ± 0.28	13.12 ± 0.04	3.85
[−21.0, −21.5]	12.20 ± 0.03	0.18 ± 0.07	11.77 ± 0.26	13.45 ± 0.03	3.84
$\log M_r$	$\log M^{\rm b}_{\rm min}$	$\sigma^{\rm b}_{\log M}$	$\log M^{\rm b}_0$	$\log M'^{\rm b}_1$	χ^2/d.o.f.
[−19.5, −20.0]	12.01 ± 0.04	0.26 ± 0.10	11.08 ± 0.33	13.27 ± 0.03	1.80
[−20.0, −20.5]	12.26 ± 0.03	0.36 ± 0.14	11.91 ± 0.10	13.46 ± 0.03	3.04
[−20.5, −21.0]	12.97 ± 0.08	0.80 ± 0.26	10.59 ± 0.42	13.81 ± 0.03	3.85
[−21.0, −21.5]	13.41 ± 0.04	0.69 ± 0.14	11.95 ± 0.23	14.41 ± 0.03	3.84

clustering analysis to the combined SDSS + *GALEX* footprint, which would discard 75 per cent of the SDSS area.[4]

Instead, we first model the SDSS correlation functions and galaxy number densities with an eight-parameter HOD described in Section 5.1.1, and then fit the colour bin sample HOD (Section 5.1.2) using the model for the SDSS samples obtained in the previous step; using the full (non-block diagonal) data covariances (Fig. 13) in each step. This method assumes that the colour sample–SDSS sample cross-correlations ($\boldsymbol{w}_{X\rm f}$, $\boldsymbol{w}_{X\rm b}$) contain little information on the HOD of the SDSS sample compared to the SDSS internal correlation functions used in the first step of the fitting procedure. This assumption is well motivated by statistical uncertainties as the colour selected samples are over an order of magnitude smaller than the SDSS reference samples. We propagate correlated uncertainties in the HOD model parameters for the SDSS reference sample to the HOD of the colour bin sample by marginalizing over 15 randomly chosen models for the SDSS HOD.

Specifically, we compute the χ^2 as

$$\chi^2 = \left(\boldsymbol{W}_Y^{\rm data} - \boldsymbol{W}_Y^{\rm model}\right) {\rm Cov}^{-1}(\boldsymbol{W}_Y) \left(\boldsymbol{W}_Y^{\rm data} - \boldsymbol{W}_Y^{\rm model}\right) + \left(\boldsymbol{n}_Y^{\rm data} - \boldsymbol{n}_Y^{\rm model}\right) {\rm Cov}^{-1}(\boldsymbol{n}_Y) \left(\boldsymbol{n}_Y^{\rm data} - \boldsymbol{n}_Y^{\rm model}\right) , \quad (25)$$

where $Y \in \{S, X\}$, with galaxy number densities $\boldsymbol{n}_S = (n_{\rm f}, n_{\rm b})$ or $\boldsymbol{n}_X = n_X$, and with the statistical error on the number densities ${\rm Cov}(\boldsymbol{n}_Y)$ estimated from field to field variations. The HOD parameter space is explored using a Markov Chain Monte Carlo method with a multivariate Gaussian proposal function and flat priors $\{\log_{10} M_{\rm min}, \log_{10} M_{\rm c}^X, \log_{10} M_1, \log_{10} M'_1\} \in [11, 17]$, $\{\sigma_{\log M}, \sigma_X\} \in [0.05, 1.0]$, $\{\alpha_X, f_X\} \in [0.5, 2.0]$ and $\log_{10} M_0 \in [8, 15]$. At each step a new set of HOD parameters is always accepted if $\chi^2_{\rm new} \leq \chi^2_{\rm old}$, and it is accepted with probability $\exp(-(\chi^2_{\rm new} - \chi^2_{\rm old})/2)$ if $\chi^2_{\rm new} > \chi^2_{\rm old}$. The typical chain length is 20 000, and we compare 10 chains of length 20 000 and one chain of length 100 000 to test convergence.

5.4 Results

Our best-fitting HOD model parameters for the SDSS samples and their marginalized 1σ errors are given in Table 4. Our results agree well with the corresponding luminosity threshold samples in the analysis of Zehavi et al. (2011), and we confirm the overall trend of characteristic halo masses for hosting central and satellite galaxies with luminosity threshold. For a detailed comparison, note that these two analyses use different fitting formulae for the halo mass function, halo bias and halo mass–concentration relations.

Based on these HOD models for the SDSS reference samples, we now turn to the colour selected galaxy samples. Fig. 15 shows the measured CCFs between colour samples and the SDSS reference samples, the best-fitting model correlation functions and the best-fitting HOD. For comparison, we also show the properly weighted sum of the colour sample HOD models, and the best-fitting HOD for the colour-independent sample of SDSS galaxies in the sample luminosity bin. While the characteristic mass scales of these HODs are similar, such comparisons are limited by the large degeneracies between fit parameters.[5] Overall, these models provide acceptable fits to the measured correlation functions, with an exception for the green and red galaxy samples in luminosity bin [−20.5, −21.0]. These correlation functions have an unusual flat shape, do not show the characteristic transition from one-halo to two-halo term regimes, and the typical host halo masses inferred from the two-halo regime are significantly larger than those inferred from the one-halo term only. As discussed in Section 4.3, the redshift of this luminosity bin is centred on the Sloan Great Wall, which is contained almost completely in the angular mask of the SDSS–*GALEX* cross-match. Hence the clustering measurements in this luminosity bin may be affected by the overdense environment, and as the great Wall occupies a disproportionally large fraction of the combined footprint, the jackknife error bars may underestimate the sample variance. For comparison, we show the CCFs of $(g - r)$ colour identified red galaxies in this luminosity bin computed over the full SDSS area and the combined survey footprint in Fig. 16. The clustering of (NUV − r) and $(g - r)$ selected red galaxies in the joint survey geometry is nearly indistinguishable, while the CCF of red galaxies in this luminosity bin over the full SDSS area has the expected shape. It can be fitted with a colour bin HOD model with reduced $\chi^2 = 3.2$, suggesting that the poor fit in Fig. 15 is indeed caused by the Great Wall structure and not a systematic effect in the construction of the (NUV − r) selected galaxy sample.

Fig. 17 and Table 5 show marginalized constraints on the mean mass of haloes hosting a central galaxy of given colour and

[4] Also note that even if one was willing to discard most of the SDSS data, obtaining an invertible joint covariance for the five different correlation functions, sampled with $N_{\rm bin}$ radial bins, would require dividing the joint footprint into more than $5N_{\rm bin}$ equal-area jackknife regions $N_{\rm sub}$. Additionally, the correction factor required to obtain an unbiased estimate of the inverse covariance scales as the ratio of the number of bins (data vector variables) to the number of data sets (Hartlap et al. 2007), resulting either in very large error bars ($N_{\rm sub} \sim 5N_{\rm bin}$) or restricting the analysis to very small scales ($N_{\rm sub} \gg 5N_{\rm bin}$).

[5] Ideally, one would fit all three colour samples simultaneously and use the sum of the three colour sample HODs to fit the correlation functions of the colour independent luminosity bin sample. However, the survey area of our current sample is not sufficient to estimate the large covariance matrices required for such an analysis.

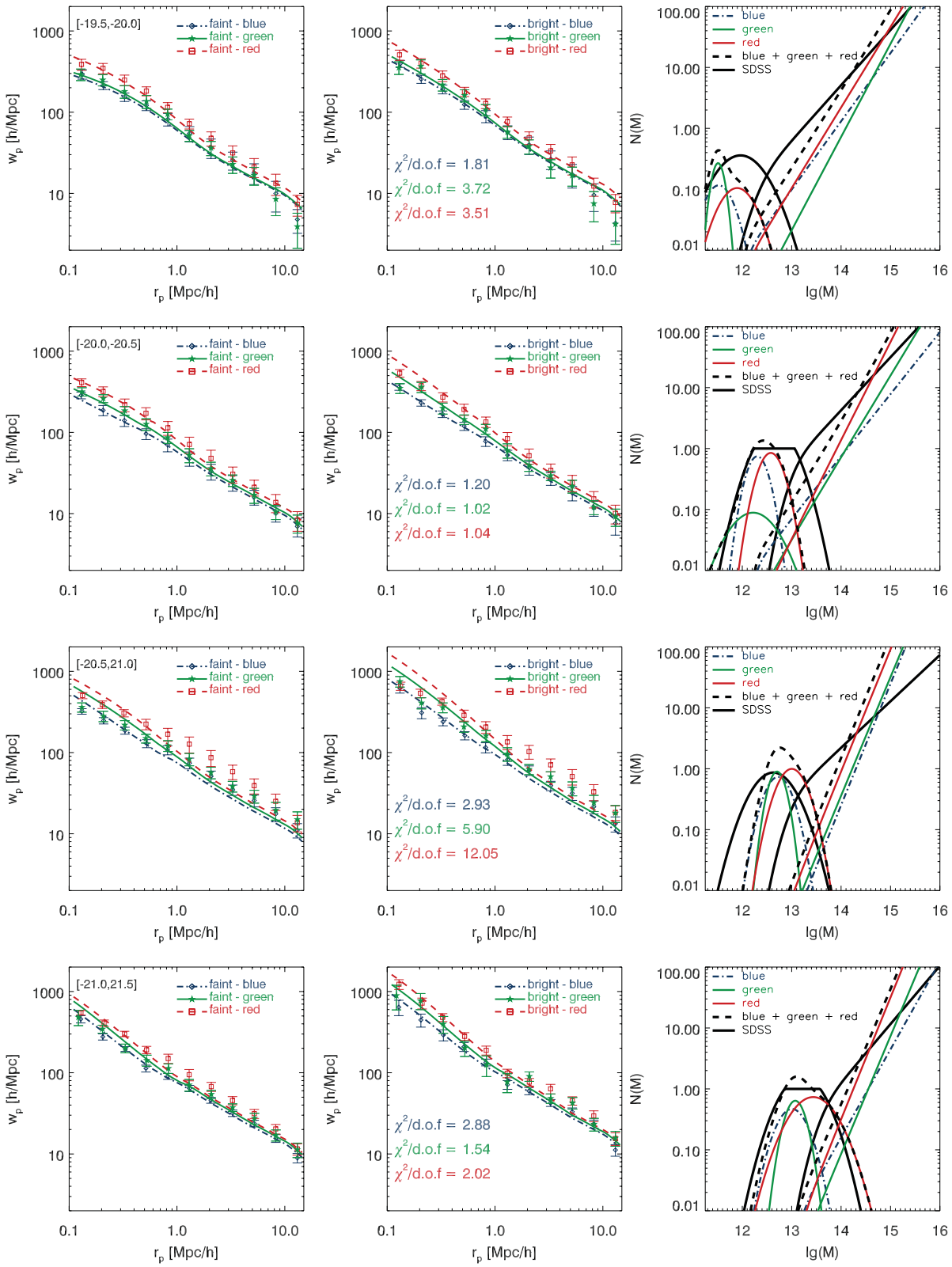

Figure 15. Each row shows the measured correlation functions and best-fitting HOD of (NUV − *r*) selected galaxy samples for one luminosity bin. The left/middle panel show the cross-correlation measurements using the faint/bright sample and their joint fit. We list the reduced χ^2 of these fits in the middle panel. The right-hand panel shows the colour sample HOD derived from fitting these CCFs, the sum of all the colour samples and the best-fitting HOD of all SDSS galaxies in the same luminosity bin.

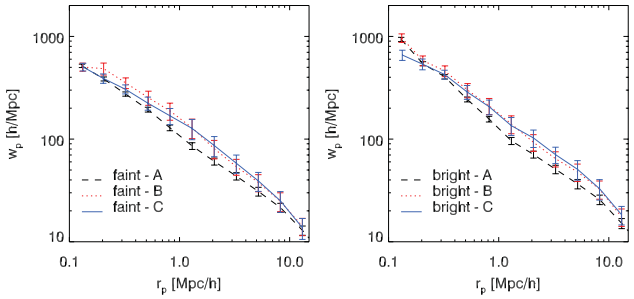

Figure 16. CCFs of red galaxies in luminosity bin [−20.5, −21.0] for different survey areas. The dashed lines are the CCFs with all $(g-r) > 0.85$ galaxies in SDSS in this magnitude bin, the dotted line restricts the SDSS red galaxies to the combined footprint, and the solid line shows the CCF for (NUV − r) identified red galaxies.

luminosity (which is different from M_c^X, as it also depends on the scatter σ_X), satellite fraction and HOD-derived galaxy bias for colour and luminosity bin samples based on the parametrization described in Section 5.1.2. We show these derived quantities instead of the HOD parameters as they are less affected by degeneracies between the fit parameters, which cause large marginalized errors in the individual fit parameters.

Based on this simple parametrization, we find red central galaxies to occupy more massive haloes than the average central galaxy from the same luminosity bin. Within the statistical uncertainty due to the small size of our colour selected galaxy samples, there is no significant difference between the halo masses of blue and green central galaxies. At fixed luminosity, the satellite fraction and HOD derived galaxy bias increases with (NUV − r) colour. The former is consistent with the results of Zehavi et al. (2011) who found the satellite fraction to vary smoothly with $(g-r)$ colour at fixed luminosity. This analysis used a one-parameter family of models based on the HOD of the colour-independent luminosity threshold sample with only the normalization of the satellite galaxy occupation function as a free parameters. Note that given the similarities in central galaxy halo masses, differences in the HOD-derived bias parameters mainly reflect the changes in the mean halo mass for satellite galaxies. This implies that the host halo masses of green satellite galaxies are intermediate between those of blue and red satellite galaxies.

Overall, we find the slope of the satellite occupation distribution, α_X, and radial concentration parameter, c_X, to increase with (NUV − r) colour. However, the degeneracies between HOD parameters are large and do not allow us to put reliable constraints on their luminosity dependence.

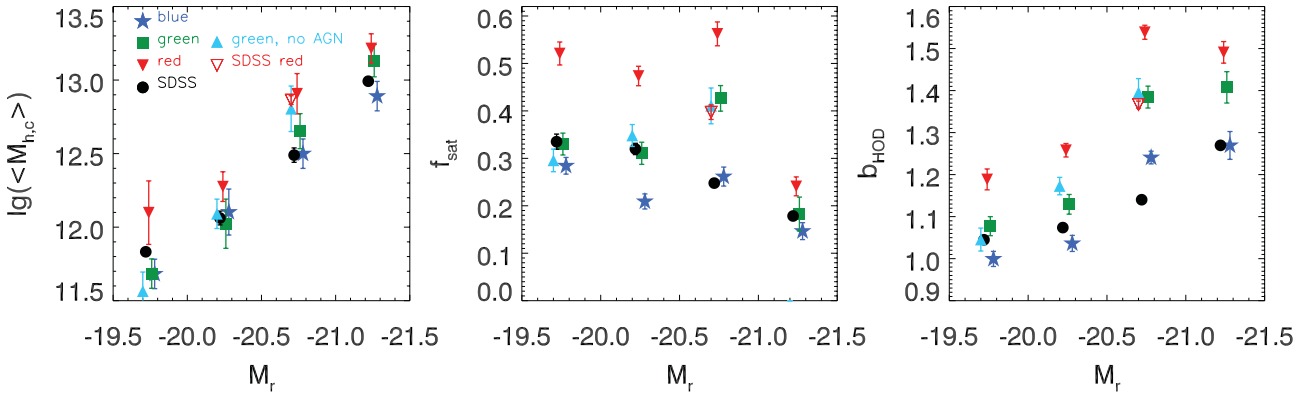

Figure 17. Derived HOD parameters for luminosity and colour bin samples. Left: mean halo mass for a halo having a central galaxy from a particular sample. Middle: satellite fraction as a function of galaxy luminosity and colour. Right: galaxy bias derived from the HOD model fit.

Table 5. Best-fitting HOD model derived parameters and their correlation coefficients, ρ, for colour selected galaxy samples.

$\log M_r$	Colour	$\log\langle M_{\rm h,c}\rangle$	$f_{\rm sat}$	$b_{\rm HOD}$	χ^2/d.o.f.	$\rho(\lg\langle M_{\rm h,c}\rangle, f_{\rm sat})$	$\rho(\lg\langle M_{\rm h,c}\rangle, b_{\rm HOD})$	$\rho(f_{\rm sat}, b_{\rm HOD})$
[−19.5, −20.0]	Blue	11.68 ± 0.10	0.28 ± 0.02	0.99 ± 0.02	1.81	−0.13	0.45	0.82
[−19.5, −20.0]	Green	11.68 ± 0.12	0.33 ± 0.03	1.07 ± 0.02	3.72	0.05	0.35	0.91
[−19.5, −20.0]	Red	12.09 ± 0.21	0.51 ± 0.03	1.18 ± 0.03	3.51	−0.36	0.58	0.54
[−19.5, −20.0]	All	11.83 ± 0.03	0.33 ± 0.01	1.04 ± 0.01	2.04	0.21	−0.01	0.09
[−20.0, −20.5]	Blue	12.10 ± 0.15	0.21 ± 0.02	1.04 ± 0.02	1.20	−0.60	0.38	0.73
[−20.0, −20.5]	Green	12.02 ± 0.17	0.31 ± 0.02	1.12 ± 0.02	1.02	−0.49	0.51	0.46
[−20.0, −20.5]	Red	12.28 ± 0.10	0.47 ± 0.02	1.25 ± 0.01	1.04	−0.39	0.11	0.86
[−20.0, −20.5]	All	12.06 ± 0.04	0.31 ± 0.01	1.07 ± 0.01	1.30	−0.73	0.41	0.26
[−20.5, −21.0]	Blue	12.49 ± 0.11	0.26 ± 0.02	1.24 ± 0.02	2.93	−0.59	0.47	0.41
[−20.5, −21.0]	Green	12.65 ± 0.12	0.42 ± 0.03	1.38 ± 0.03	5.90	−0.34	0.46	0.67
[−20.5, −21.0]	Red	12.91 ± 0.14	0.56 ± 0.03	1.53 ± 0.02	12.06	−0.69	0.35	0.42
[−20.5, −21.0]	All	12.49 ± 0.05	0.25 ± 0.01	1.14 ± 0.01	2.32	−0.52	0.77	−0.01
[−21.0, −21.5]	Blue	12.89 ± 0.09	0.14 ± 0.01	1.26 ± 0.02	2.88	−0.69	0.81	−0.24
[−21.0, −21.5]	Green	13.13 ± 0.11	0.18 ± 0.03	1.41 ± 0.04	1.54	−0.71	0.72	−0.08
[−21.0, −21.5]	Red	13.21 ± 0.10	0.24 ± 0.02	1.49 ± 0.03	2.02	−0.56	0.67	0.13
[−21.0, −21.5]	All	12.99 ± 0.02	0.18 ± 0.01	1.27 ± 0.01	2.54	−0.65	0.79	−0.12

For luminosity bin [−20.5, −21.0], we also show results derived from $(g - r)$ selected red galaxies in the full SDSS area to indicate the impact of the Sloan Great Wall. In the Great Wall, the satellite fraction and halo mass of red galaxies are increased compared to a more representative survey volume, as expected from the colour–density relation. As the (NUV − r) colour selected samples in this luminosity bin are subject to increased sample variance, the results for blue and green galaxies in this luminosity bin should similarly be interpreted with caution.

As noted by Martin et al. (2007) and Salim et al. (2007), a large fraction of AGNs have green (NUV − r) colours. These galaxies may be transitional galaxies with star formation being quenched by AGN feedback (e.g. after undergoing a major merger; Springel, Di Matteo & Hernquist 2005), or red sequence interlopers which appear green due to the NUV AGN continuum emission. We test whether the intermediate clustering of green valley galaxies is caused by AGN, which may be a different population from the non-AGN transitional galaxies. We identify green AGN through emission line diagrams (Baldwin et al. 1981) using the Kewley et al. (2001) extreme starburst classification line. We use the emission line measurements from the MPA-JHU catalogue and require a S/N $\geq$ 3 in the emission lines. Our goal is to remove any potential AGN contamination from the green valley galaxy sample, and we remove all galaxies which are classified as AGNs in at least one of the three diagrams, as this allows us to categorize galaxies which do not meet the S/N threshold for all emission line. Repeating our clustering and HOD analysis for non-AGN green galaxies, we find the HOD of green non-AGN galaxies to be indistinguishable from that of green galaxies including AGN, in agreement with trends earlier observed by Li et al. (2006) and Heinis et al. (2009). We do not show results derived from HOD fits for the non-AGN green valley galaxies in luminosity bin [−21.0, −21.5] as this sample is too small to obtain stable covariances.

6 SUMMARY AND DISCUSSION

We introduced a new analysis and HOD modelling technique for galaxy CCFs using multiple tracer populations. This approach is particularly useful for interpreting the clustering of sparse and/or luminosity bin selected galaxy samples of interest. It is advantageous for the analysis of sparse galaxy samples as considering the CCF with more abundant galaxy populations significantly reduces the statistical uncertainty.

While the galaxy number density provides strong constraints on the HOD of luminosity threshold samples, the HOD of luminosity bin samples is independent of the galaxy abundance; in this case considering the cross-correlation with multiple tracer populations is particularly useful as it provides an additional mass scale for the calibration of the luminosity bin HOD. An additional advantage of this method is that modelling the CCF between a colour selected sample and a colour-independent sample does not require assumptions on the correlation between central and satellite colours.

This allows us to constrain the central galaxy HOD of colour and luminosity bin selected samples for the first time. We apply this multiple tracer technique to analyse the clustering of (NUV − r) colour selected blue, red and green valley galaxy samples. Our key result is that halo mass of central galaxies, satellite fraction and halo mass of satellite galaxies increase with (NUV − r) colour at fixed luminosity.

While our results indicate that the clustering properties of green valley galaxies are consistent with them being an intermediate population between blue and red galaxies, the (NUV − r) selected green valley galaxy samples in this analysis consist of only about 1000 galaxies and are too small to provide insight on the transition mechanism(s) at work. In particular, the HOD parameters which describe the abundance and distribution of colour selected satellite galaxies, i.e. the slope of the satellite occupation function, α_X, and the colour dependence of the radial satellite distribution, c_X, are poorly constrained by the data. With data from future galaxy redshift surveys, these parameters will provide information on the efficiency of star formation quenching as a function of halo mass and location within a halo. Furthermore, larger data sets will enable a detailed measurement of the redshift space correlation function and thus enable constraints on the infall stage and satellite orbits of transitional galaxies.

The reduced χ^2 values of the best-fitting HODs in our analysis of colour selected galaxy samples are relatively large, and our model is particularly insufficient to reproduce the clustering of galaxies in or near the Sloan Great Wall. Overall, it is not surprising that a five-parameter HOD model does not fully describe the colour-dependent clustering of galaxies. While the HOD formalism works well to describe the overall relation between (colour-independent) galaxies and their haloes, it is questionable if the strong assumptions implicit in the HOD formalism, such as the one-to-one relation between halo mass and bias, hold for each subpopulation. Additionally, numerical and observational results indicate that the influence of massive haloes may extend beyond R_{200}, e.g. through highly eccentric satellite orbits (Benson 2005; Wetzel, Tinker & Conroy 2012) and infall-related shocks extending beyond the virial radius (e.g. Balogh, Navarro & Morris 2000), which is not easily incorporated in halo models.

Finally, we note that Behroozi, Conroy & Wechsler (2010) and Leauthaud et al. (2011) recently proposed an improved HOD parametrization based on a detailed model for the relation between stellar mass and halo mass. Their results (fig. 3 in Leauthaud et al. 2011) indicate that halo masses derived from the HOD parametrization for luminosity threshold samples adopted in our analysis (equation 6) may be biased by up to 40 per cent, with the main source of this discrepancy being the assumptions of a power-law form and constant scatter for the luminosity–halo mass relation. For luminosity bin samples, however, these assumptions are better justified, and we expect only small discrepancies between different HOD parametrizations.

ACKNOWLEDGMENTS

We thank David Weinberg and Zheng Zheng for helpful discussions about HOD modelling of binned samples.

During the preparation of this work, EK and CMH were supported by the US National Science Foundation (AST-0807337) and the David & Lucile Packard Foundation. CH was additionally supported by the US Department of Energy (DE-SC0006624). Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan

Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Balogh M. L., Navarro J. F., Morris S. L., 2000, ApJ, 540, 113
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Benson A. J., 2005, MNRAS, 358, 551
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Bhattacharya S., Habib S., Heitmann K., 2011, preprint (arXiv 1112.5479)
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Blanton M. R. et al., 2003a, AJ, 125, 2348
Blanton M. R., Lin H., Lupton R. H., Maley F. M., Young N., Zehavi I., Loveday J., 2003b, AJ, 125, 2276
Blanton M. R. et al., 2005, AJ, 129, 2562
Cooper M. C. et al., 2006, MNRAS, 370, 198
Cooper M. C. et al., 2007, MNRAS, 376, 1445
Coupon J. et al., 2012, A&A, 542, A5
Eifler T., Kilbinger M., Schneider P., 2008, A&A, 482, 9
Eisenstein D. J. et al., 2001, AJ, 122, 2267
Faber S. M. et al., 2007, ApJ, 665, 265
Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, AJ, 111, 1748
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gunn J. E. et al., 1998, AJ, 116, 3040
Gunn J. E. et al., 2006, AJ, 131, 2332
Guo Q., Cole S., Eke V., Frenk C., 2012, MNRAS, 427, 428
Hamilton A. J. S., Tegmark M., 2004, MNRAS, 349, 115
Hartlap J., Simon P., Schneider P., 2007, A&A, 464, 399
Heinis S. et al., 2007, ApJS, 173, 503
Heinis S. et al., 2009, ApJ, 698, 1838
Hogg D. W., Finkbeiner D. P., Schlegel D. J., Gunn J. E., 2001, AJ, 122, 2129
Ivezić Ž. et al., 2004, Astron. Nachr., 325, 583
Kauffmann G. et al., 2003a, MNRAS, 346, 1055
Kauffmann G. et al., 2003b, MNRAS, 341, 33
Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121
Krumpe M., Miyaji T., Coil A. L., 2010, ApJ, 713, 558
Landy S. D., Szalay A. S., 1993, ApJ, 412, 64
Leauthaud A., Tinker J., Behroozi P. S., Busha M. T., Wechsler R., 2011, ApJ, 738, 45
Li C., Kauffmann G., Wang L., White S. D. M., Heckman T. M., Jing Y. P., 2006, MNRAS, 373, 457
Loh Y. et al., 2010, MNRAS, 407, 55
Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., Yasuda N., 2001, in Harnden F. R. Jr, Primini F. A., Payne H. E., eds, ASP Conf. Ser. Vol. 238, Astronomical Data Analysis Software and Systems X. Astron. Soc. Pac., San Francisco, p. 269
Martin D. C. et al., 2007, ApJS, 173, 342
Miyaji T., Krumpe M., Coil A. L., Aceves H., 2011, ApJ, 726, 83
Moran S. M., Ellis R. S., Treu T., Smith G. P., Rich R. M., Smail I., 2007, ApJ, 671, 1503
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Padmanabhan N. et al., 2008, ApJ, 674, 1217
Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, AJ, 125, 1559
Salim S. et al., 2007, ApJS, 173, 267
Simon P., Hetterscheidt M., Wolf C., Meisenheimer K., Hildebrandt H., Schneider P., Schirmer M., Erben T., 2009, MNRAS, 398, 807
Skibba R. A., Sheth R. K., 2009, MNRAS, 392, 1080
Smith J. A. et al., 2002, AJ, 123, 2121
Springel V., Di Matteo T., Hernquist L., 2005, ApJ, 620, L79
Strauss M. A. et al., 2002, AJ, 124, 1810
Swanson M. E. C., Tegmark M., Hamilton A. J. S., Hill J. C., 2008, MNRAS, 387, 1391
Szapudi S., Szalay A. S., 1998, ApJ, 494, L41
Tinker J. L., Weinberg D. H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878
Tinker J., Wetzel A., Conroy C., 2011, preprint (arXiv 1107.5046)
Treu T., Ellis R. S., Kneib J., Dressler A., Smail I., Czoske O., Oemler A., Natarajan P., 2003, ApJ, 591, 53
Tucker D. L. et al., 2006, Astron. Nachr., 327, 821
von der Linden A., Wild V., Kauffmann G., White S. D. M., Weinmann S., 2010, MNRAS, 404, 1231
Watson D. F., Berlind A. A., McBride C. K., Hogg D. W., Jiang T., 2012, ApJ, 749, 83
Wetzel A. R., Tinker J. L., Conroy C., 2012, MNRAS, 424, 232
Wyder T. K. et al., 2007, ApJS, 173, 293
York D. G. et al., 2000, AJ, 120, 1579
Zehavi I. et al., 2005, ApJ, 630, 1
Zehavi I. et al., 2011, ApJ, 736, 59
Zheng Z. et al., 2005, ApJ, 633, 791

This paper has been typeset from a TEX/LATEX file prepared by the author.